



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 145331

TO: Janet Epps-Ford
Location: rem/2c05/2c18
Art Unit: 1635
Tuesday, March 01, 2005

Case Serial Number: 10/086062

From: Barb O'Brien
Location: Biotech-Chem Library
Remsen 1A69
Phone: 571-272-2518
barbara.obrien @uspto.gov

Search Notes

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 28, 2005, 23:32:36 ; Search time 729.221 Seconds
(without alignments)
1661.198 Million cell updates/sec

Title: US-10-086-062-1

Perfect score: 25
Sequence: 1 ctgagaccctctcgaagttccgct 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_srb:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	25	6	AX358106 Sequence
2	25	100.0	813	6	AR097833 Sequence
3	25	100.0	994	6	AR097835 Sequence
4	25	100.0	1807	6	AR097836 Sequence
5	25	100.0	2182	8	AY342393 Zea diplo
6	25	100.0	2620	6	AR097837 Sequence
7	25	100.0	2807	6	AX134224 Sequence
8	25	100.0	2949	6	AX259248 Sequence
9	25	100.0	3069	6	AX202417 Sequence
10	25	100.0	3121	6	AX202416 Sequence
11	25	100.0	3433	6	AR097838 Sequence
12	25	100.0	3840	6	I06936 Sequence 8
13	25	100.0	3840	6	I38571 Sequence 1
14	25	100.0	3840	6	AR365568 Sequence
15	25	100.0	3841	8	S94464 polynubiquit
16	25	100.0	3972	6	AX355352 Sequence
17	25	100.0	4190	8	ZMU29159 Zea maye c1
18	25	100.0	4341	6	AX259249 Sequence
19	25	100.0	5465	6	BD194459 Chimeric

20	25	100.0	5522	12	AY452753 Reporter
21	25	100.0	5622	6	AX039929 Sequence
22	25	100.0	6103	6	BD194458 Chimeric
23	25	100.0	6111	6	BD247866 Transgent
24	25	100.0	6111	6	AR193435 Sequence
25	25	100.0	6111	6	AR430220 Sequence
26	25	100.0	6215	6	AX118823 Sequence
27	25	100.0	6217	6	AR209726 Sequence
28	25	100.0	6579	6	AX118824 Sequence
29	25	100.0	6667	6	AX118825 Sequence
30	25	100.0	6818	6	AX840287 Sequence
31	25	100.0	6898	6	BD132717 Regulator
32	25	100.0	6898	6	AR209020 Sequence
33	25	100.0	6898	6	AR482089 Sequence
34	25	100.0	7018	12	AY452736 Reporter
35	25	100.0	7027	6	AX840286 Sequence
36	25	100.0	9172	6	AX259246 Sequence
37	25	100.0	9359	6	AX384394 Sequence
38	25	100.0	9359	6	AX473364 Sequence
39	25	100.0	11643	6	AX840289 Sequence
40	25	100.0	12817	6	AX356864 Sequence
41	25	100.0	12849	6	BD247868 Transgent
42	25	100.0	12949	6	AR193437 Sequence
43	25	100.0	12949	6	AR430222 Sequence
44	25	100.0	12982	6	BD091582 Vector fo
45	25	100.0	12982	6	BD091582 Vector fo

ALIGNMENTS

RESULT 1	AX358106	Sequence 1 from Patent WO0194394.	25 bp	DNA	linear	PAT 13-FEB-2002
LOCUS	AX358106					
DEFINITION	AX358106					
ACCESSION	AX358106.1	GI:18674853				
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
FEATURES						
source						
ORIGIN						
Query Match						
Best Local Similarity						
Matches						
Qy						
Db						
RESULT 2						
LOCUS						
DEFINITION						
ACCESSION						
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
Unclassified.						

REFERENCE 1 (bases 1 to 813)
AUTHORS Bowen,B.A., Bruce,W.B., Lu,G., Sims,L.E. and Tagliani,L.A.
TITLE Synthetic promoters
JOURNAL Patent: US 6072050-A 13 06-JUN-2000;
FEATURES Location/Qualifiers
source 1..813
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 813;
Best Local Similarity 100.0%; Pred.No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGACCCCTCTCGAGAGTTCGCT 25
Db 654 CTGACCCCTCTCGAGAGTTCGCT 678

RESULT 3
LOCUS AR097835 994 bp DNA linear PAT 14-FEB-2001
DEFINITION Sequence 15 from patent US 6072050.
ACCESSION AR097835
VERSION AR097835.1 GI:12806565
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 994)
AUTHORS Bowen,B.A., Bruce,W.B., Lu,G., Sims,L.E. and Tagliani,L.A.
TITLE Synthetic promoters
JOURNAL Patent: US 6072050-A 15 06-JUN-2000;
FEATURES Location/Qualifiers
source 1..994
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 994;
Best Local Similarity 100.0%; Pred.No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGACCCCTCTCGAGAGTTCGCT 25
Db 654 CTGACCCCTCTCGAGAGTTCGCT 678

RESULT 4
LOCUS AR097836 1807 bp DNA linear PAT 14-FEB-2001
DEFINITION Sequence 16 from patent US 6072050.
ACCESSION AR097836
VERSION AR097836.1 GI:12806566
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1807)
AUTHORS Bowen,B.A., Bruce,W.B., Lu,G., Sims,L.E. and Tagliani,L.A.
TITLE Synthetic promoters
JOURNAL Patent: US 6072050-A 16 06-JUN-2000;
FEATURES Location/Qualifiers
source 1..1807
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 1807;
Best Local Similarity 100.0%; Pred.No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGACCCCTCTCGAGAGTTCGCT 25

Db 654 CTGACCCCTCTCGAGAGTTCGCT 678

RESULT 5
LOCUS AY342393 2182 bp DNA linear PLN 01-AUG-2004
DEFINITION Zea diploperennis polyubiquitin-1 (Ubi-1) gene, promoter region and 5' UTR.
ACCESSION AY342393
VERSION AY342393.1 GI:37912415
KEYWORDS
SOURCE Zea diploperennis
ORGANISM Zea diploperennis
REFERENCE 1 (bases 1 to 2182)
AUTHORS Bukaryoka; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
TITLE 1 (bases 1 to 2182)
Streatfield,S.J. and Love,R.T.
Analysis of the maize polyubiquitin-1 promoter heat shock elements and generation of promoter variants with modified expression characteristics
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2182)
AUTHORS Streatfield,S.J. and Love,R.T.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-2003) Molecular Biology, ProdiGene, 101 Gateway Boulevard, College Station, TX 77845, USA
FEATURES Location/Qualifiers
source 1..2182
/organism="Zea diploperennis"
/mol_type="genomic DNA"
/db_xref="taxon:4576"
1..>2182
/gene="Ubi-1"
1..922
/gene="Ubi-1"
923..>1003
/gene="Ubi-1"
/product="polyubiquitin-1"
923..1003
/gene="Ubi-1"

ORIGIN
Query Match 100.0%; Score 25; DB 8; Length 2182;
Best Local Similarity 100.0%; Pred.No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGACCCCTCTCGAGAGTTCGCT 25
Db 708 CTGACCCCTCTCGAGAGTTCGCT 732

RESULT 6
LOCUS AR097837 2620 bp DNA linear PAT 14-FEB-2001
DEFINITION Sequence 17 from patent US 6072050.
ACCESSION AR097837
VERSION AR097837.1 GI:12806567
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2620)
AUTHORS Bowen,B.A., Bruce,W.B., Lu,G., Sims,L.E. and Tagliani,L.A.
TITLE Synthetic promoters
JOURNAL Patent: US 6072050-A 17 06-JUN-2000;
FEATURES Location/Qualifiers
source 1..2620
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 2620;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTGACCCCTCTCGAGAGTTCGGCT 25
DB 654 CTGACCCCTCTCGAGAGTTCGGCT 678

RESULT 7
AX134224 2807 bp DNA linear PAT 29-MAY-2001
LOCUS AX134224
DEFINITION Sequence 1 from Patent WO0132897.
ACCESSION AX134224
VERSION AX134224.1 GI:14270967
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.

REFERENCE 1
AUTHORS Groenewald, J.H. and Botha, F.C.
TITLE A high level, stable, constitutive promoter element for plants
JOURNAL Patent: WO 0132897-A 1 10-MAY-2001;
South African Sugar Association (ZA) ; University of Stellenbosch,
Institute for Plant Biotechnology (ZA)

FEATURES
source Location/Qualifiers
1..2807
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="synthetic construct"

ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 2807;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTGACCCCTCTCGAGAGTTCGGCT 25
DB 1501 CTGACCCCTCTCGAGAGTTCGGCT 1525

RESULT 8
AX259248 2949 bp DNA linear PAT 26-OCT-2001
LOCUS AX259248
DEFINITION Sequence 46 from Patent WO0173087.
ACCESSION AX259248
VERSION AX259248.1 GI:16508494
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.

REFERENCE 1
AUTHORS Hohn, T., Stavolone, L., de Haan, P.T., Ligon, H.T. and Kononova, M.
TITLE Cestrum yellow leaf curling virus promoters
JOURNAL Patent: WO 0173087-A 46 04-OCT-2001;
Syngenta Participations AG (CH)

FEATURES
source Location/Qualifiers
1..2949
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="artificial sequence ZmUbi-GFP-35S term expression cassette"

ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 2949;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTGACCCCTCTCGAGAGTTCGGCT 25
DB 692 CTGACCCCTCTCGAGAGTTCGGCT 716

RESULT 9
AX202417 3069 bp DNA linear PAT 30-AUG-2001
LOCUS AX202417
DEFINITION Sequence 5 from Patent WO0152620.
ACCESSION AX202417
VERSION AX202417.1 GI:15392164
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.

REFERENCE 1
AUTHORS Barbash, C.F., Stege, J.T., Guan, X. and Daimia, B.
TITLE Methods and compositions to modulate expression in plants
JOURNAL Patent: WO 0152620-A 5 26-JUL-2001;
The Scripps Research Institute (US) ; SYNGENTA AGRICULTURAL
DISCOVERY, INC. (CA)

FEATURES
source Location/Qualifiers
1..3069
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="PND3018 coding region"

ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 3069;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTGACCCCTCTCGAGAGTTCGGCT 25
DB 677 CTGACCCCTCTCGAGAGTTCGGCT 701

RESULT 10
AX202416 3121 bp DNA linear PAT 30-AUG-2001
LOCUS AX202416
DEFINITION Sequence 4 from Patent WO0152620.
ACCESSION AX202416
VERSION AX202416.1 GI:15392163
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.

REFERENCE 1
AUTHORS Barbash, C.F., Stege, J.T., Guan, X. and Daimia, B.
TITLE Methods and compositions to modulate expression in plants
JOURNAL Patent: WO 0152620-A 4 26-JUL-2001;
The Scripps Research Institute (US) ; SYNGENTA AGRICULTURAL
DISCOVERY, INC. (CA)

FEATURES
source Location/Qualifiers
1..3121
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="PND3008 coding region"

ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 3121;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTGACCCCTCTCGAGAGTTCGGCT 25
DB 677 CTGACCCCTCTCGAGAGTTCGGCT 701

RESULT 11
AR097838 3433 bp DNA linear PAT 14-FEB-2001
LOCUS AR097838
DEFINITION Sequence 18 from patent US 6072050.
ACCESSION AR097838

VERSION AR097838.1 GI:12806568
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 3433)
TITLE Bowen,B.A., Bruce,W.B., Lu,G., Sims,L.E. and Tagliani,L.A.
JOURNAL Synthetic promoters
PATENT: US 6072050-A 18 06-JUN-2000;
FEATURES Location/Qualifiers
source 1..3433
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 3433;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTGAGCCCTCTCGAGAGTCCGCT 25
Db 654 CTGAGCCCTCTCGAGAGTCCGCT 678

RESULT 12
LOCUS 106936 3840 bp DNA linear PAT 02-DEC-1994
DEFINITION Sequence 8 from Patent EP 0342926.
ACCESSION 106936
VERSION 106936.1 GI:589816
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 3840)
Quail,P.H., Christiansen,A.H., Hershey,H.P., Sharrock,R.A. and Sullivan,T.D.
TITLE Plant ubiquitin promoter system
JOURNAL Patent: EP 0342926-A2 8 23-NOV-1989;
FEATURES Location/Qualifiers
source 1..3840
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 3840;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTGAGCCCTCTCGAGAGTCCGCT 25
Db 686 CTGAGCCCTCTCGAGAGTCCGCT 710

RESULT 13
LOCUS 138571 3840 bp DNA linear PAT 13-MAY-1997
DEFINITION Sequence 1 from patent US 5614399.
ACCESSION 138571
VERSION 138571.1 GI:2084625
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 3840)
Quail,P.H., Christiansen,A.H., Hershey,H.P., Sharrock,R.A. and Sullivan,T.D.
TITLE Plant ubiquitin promoter system
JOURNAL Patent: US 5614399-A 1 25-MAR-1997;
FEATURES Location/Qualifiers
source 1..3840
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 3840;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTGAGCCCTCTCGAGAGTCCGCT 25
Db 686 CTGAGCCCTCTCGAGAGTCCGCT 710

RESULT 14
LOCUS AR365568 3840 bp DNA linear PAT 03-SEP-2003
DEFINITION Sequence 1 from patent US 5510474.
ACCESSION AR365568
VERSION AR365568.1 GI:34429351
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 3840)
Quail,P.H., Christiansen,A.H., Hershey,H.P., Sharrock,R.A. and Sullivan,T.D.
TITLE Plant ubiquitin promoter system
JOURNAL Patent: US 5510474-A 1 23-APR-1996;
FEATURES Location/Qualifiers
source 1..3840
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 3840;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTGAGCCCTCTCGAGAGTCCGCT 25
Db 686 CTGAGCCCTCTCGAGAGTCCGCT 710

RESULT 15
LOCUS S94464 3841 bp DNA linear PLN 07-MAY-1993
DEFINITION polyubiquitin [maize, Genomic, 3841 nt].
ACCESSION S94464
VERSION S94464.1 GI:248336
KEYWORDS
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 3841)
Christensen,A.H., Sharrock,R.A. and Quail,P.H.
TITLE Maize polyubiquitin genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation
JOURNAL Plant Mol. Biol. 18 (4), 675-689 (1992)
MEDLINE 92216044
PUBMED 1313711
REMARK GenBank staff at the National Library of Medicine created this entry [NCBI gidsq 94464] from the original journal article.
FEATURES Location/Qualifiers
source 1..3841
/organism="Zea mays"
/mol_type="genomic DNA"
/db_xref="taxon:4577"
1..3841
/gene="polyubiquitin"
1994..3595
/gene="polyubiquitin"
/codon_start=1

```

/protein_id="AAB21993.1"
/db_xref="GI:248337"
/translation="MQIFVKLTIGKTTITLEVESSDTIDNVKAKIQDKEGIPPDQORLI
FAGKQLEDGRTLDVNIQKESTIHLVLRGGMQIFVKLTIGKTTITLEVESSDTIDN
KAKIQDKEGIPPDQORLI FAGKQLEDGRTLDVNIQKESTIHLVLRGGMQIFVKLT
IGKTTITLEVESSDTIDNVKAKIQDKEGIPPDQORLI FAGKQLEDGRTLDVNIQKEST
IHLVLRGGMQIFVKLTIGKTTITLEVESSDTIDNVKAKIQDKEGIPPDQORLI FAGK
QLEDGRTLDVNIQKESTIHLVLRGGMQIFVKLTIGKTTITLEVESSDTIDNVKAKI
QDKEGIPPDQORLI FAGKQLEDGRTLDVNIQKESTIHLVLRGGMQIFVKLTIGKT
ITLEVESSDTIDNVKAKIQDKEGIPPDQORLI FAGKQLEDGRTLDVNIQKESTIHLV
LRGGMQIFVKLTIGKTTITLEVESSDTIDNVKAKIQDKEGIPPDQORLI FAGKQLED
GRTLDVNIQKESTIHLVLRGGMQ"

```

ORIGIN

```

Query Match      100.0%; Score 25; DB 8; Length 384;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 CTGACCCCTCTCGAGAGTTCGCT 25
      |||||
Db      687 CTGACCCCTCTCGAGAGTTCGCT 711

```

Search completed: March 1, 2005, 00:29:10
 Job time : 735.221 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 28, 2005, 23:26:41 ; Search time 218.182 Seconds
(without alignments)
678.304 Million cell updates/sec

Title: US-10-086-062-1

Perfect score: 25
Sequence: 1 ctgacccctctcgaggttcgcgc 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1Dec04:*

- 1: geneseqn19808:*
- 2: geneseqn19908:*
- 3: geneseqn20008:*
- 4: geneseqn2001as:*
- 5: geneseqn2001bs:*
- 6: geneseqn2002as:*
- 7: geneseqn2002bs:*
- 8: geneseqn2003as:*
- 9: geneseqn2003bs:*
- 10: geneseqn2003cs:*
- 11: geneseqn2004as:*
- 12: geneseqn2004bs:*
- 13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	25	100.0	25	AAD24066
2	25	100.0	813	AAZ11619
3	25	100.0	813	AAZ11619
4	25	100.0	994	AAZ11621
5	25	100.0	1413	AAZ11621
6	25	100.0	1807	AAZ11622
7	25	100.0	2005	AAZ11622
8	25	100.0	2100	AAZ11622
9	25	100.0	2609	AAZ11622
10	25	100.0	2620	AAZ11623
11	25	100.0	2807	AAH19412
12	25	100.0	2848	AAH19412
13	25	100.0	3069	AAH19412
14	25	100.0	3121	AAH19412
15	25	100.0	3433	AAH19412
16	25	100.0	3840	AAH19412
17	25	100.0	3972	AAH19412
18	25	100.0	4341	AAH19412
19	25	100.0	4341	AAH19412
20	25	100.0	4569	AAH19412

21	25	100.0	4587	9	ACF06316	ACF06316	Intact v1
22	25	100.0	5465	2	AAZ20088	AAZ20088	Plasmid p
23	25	100.0	5622	3	AAZ0506	AAZ0506	DNA cont
24	25	100.0	6103	2	AAZ20087	AAZ20087	Plasmid p
25	25	100.0	6111	3	AAZ54211	AAZ54211	Transform
26	25	100.0	6215	4	AAZ03876	AAZ03876	Noti frag
27	25	100.0	6579	4	AAZ03877	AAZ03877	Noti frag
28	25	100.0	6667	4	AAZ03878	AAZ03878	Noti frag
29	25	100.0	6818	10	ACF58342	ACF58342	Nucleotid
30	25	100.0	6898	2	AAV63742	AAV63742	Plasmid U
31	25	100.0	7027	10	ACF58341	ACF58341	Nucleotid
32	25	100.0	8352	12	ADP73862	ADP73862	Plasmid p
33	25	100.0	9144	12	ADP73861	ADP73861	Plasmid p
34	25	100.0	9172	4	AAZ19836	AAZ19836	Plasmid p
35	25	100.0	9359	6	ABK11039	ABK11039	PVDH36 v
36	25	100.0	9359	6	ABK10687	ABK10687	Transform
37	25	100.0	11643	10	ACF58344	ACF58344	Nucleotid
38	25	100.0	12380	12	ADP73863	ADP73863	Plasmid p
39	25	100.0	12817	6	ABK15668	ABK15668	Expressio
40	25	100.0	12949	3	AAZ54213	AAZ54213	Transform
41	25	100.0	12982	4	AAZ6432	AAZ6432	Vector ps
42	25	100.0	12982	4	AAZ6432	AAZ6432	Vector ps
43	25	100.0	13269	6	AAZ6361	AAZ6361	Maize ubi
44	25	100.0	13274	6	ABK15666	ABK15666	Binary ve
45	25	100.0	13829	11	ADN08781	ADN08781	Maize opt

ALIGNMENTS

RESULT 1	AAZ24066	standard; DNA; 25 BP.
ID	AAZ24066	
XX	AAZ24066;	
AC	AAZ24066;	
XX	AAZ24066;	
DT	09-APR-2002	(first entry)
XX	09-APR-2002	
DE	Maize wild-type Ubi-1 promoter heat shock element.	
XX	Maize wild-type Ubi-1 promoter heat shock element.	
KW	Gene expression; maize; ubiquitin promoter; Ubi-1; HSE;	
KW	heat shock element; agronomic gene; ds.	
XX	heat shock element; agronomic gene; ds.	
OS	Zea mays.	
XX	Zea mays.	
FT	Key	Location/Qualifiers
FT	misc_feature	1..15
FT	misc_feature	/tag= "a"
FT	misc_feature	/note= "5' heat shock element"
FT	misc_feature	/tag= "b"
FT	misc_feature	/note= "3' heat shock element"
XX	misc_feature	
PN	WO200194394-A2.	
XX	WO200194394-A2.	
PD	13-DEC-2001.	
XX	13-DEC-2001.	
PF	08-JUN-2001; 2001WO-US018689.	
XX	08-JUN-2001; 2001WO-US018689.	
PR	09-JUN-2000; 2000US-00590558.	
XX	09-JUN-2000; 2000US-00590558.	
PA	(PROD-) PRODIGENE INC.	
XX	(PROD-) PRODIGENE INC.	
PI	Jilka JM, Hood BE, Howard JA;	
XX	Jilka JM, Hood BE, Howard JA;	
DR	WPI; 2002-122117/16.	
XX	WPI; 2002-122117/16.	
PT	New promoter sequences for causing expression of a structural gene	
PT	especially agronomic gene or open reading frame in a plant cell,	
XX	comprises engineered versions of the maize ubiquitin promoter.	
PS	Example 1; Page 43; 68p; English.	
XX	Example 1; Page 43; 68p; English.	

CC The invention relates to a promoter sequence capable of directing
CC expression of a nucleotide sequence in a plant cell, comprising maize
CC ubiquitin (Ubi-1) promoter sequence with a modification so that it does
CC not include two overlapping heat shock elements (HSE) or it directs
CC expression to increase the endosperm/embryo expression ratio of the
CC protein when compared to the ratio from a wild-type ubiquitin promoter.
CC The modified Ubi-1 promoter comprises a deletion of 3', 5' or both HSEs,
CC two non-overlapping/adjacent HSEs, replacement of HSEs with a trimer of a
CC seed specific element from the promoter of pea lectin gene pL1, or
CC insertion of a transcription factor binding site in the HSE region. An
CC expression construct comprising modified Ubi-1 promoter is useful for
CC causing expression of a structural gene (agronomic genes) or open reading
CC frame in a plant cell. The modified Ubi-1 promoter increases expression
CC levels beyond those observed with native ubiquitin promoter. The present
CC sequence is maize wild-type Ubi-1 promoter heat shock element
XX

SQL Sequence 25 BP, 3 A; 10 C; 6 G; 6 T; 0 U; 0 Other;

Query Match 100.0%; Score 25; DB 6; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTGAGCCCTCTCGAGATTCCGCT 25
DB 1 CTGAGCCCTCTCGAGATTCCGCT 25

RESULT 2
AAZ11619
ID AAZ11619 standard; DNA; 813 BP.
XX
AC AAZ11619;
XX
DT 16-NOV-1999 (first entry)
XX

DE Maize Ubi1 upstream activating region (UAR).
XX
XX Plant promoter; TATA motif; transcription start site; upstream element;
XX gene expression; oxalate oxidase; plant resistance; pathogen; maize;
XX Ubi-1 promoter; upstream activating region; ss.

OS Zea sp.
XX
XX WO9943838-A1.
XX
XX 02-SEP-1999.
XX
XX 23-FEB-1999; 99WO-US003863.
XX
XX 24-FEB-1998; 98US-00028819.
XX

PA (PION-) PIONEER HI-BRED INT INC.
XX

PI Bowen BA, Bruce WB, Lu G, Sims LE, Tagliani LA;
XX
XX WPI; 1999-540601/45.
XX

PT New synthetic promoter functional in plants to provide non-tissue
PT specific, constitutive expression, particularly of oxalate oxidase for
PT increased resistance to pathogens.
XX
XX Disclosure; Page 51; 61pp; English.

XX The invention provides a new synthetic plant promoter that comprises a
CC TATA motif; a transcription start site (TSS) and a region between TATA
CC and TSS containing at least 64 percent GC content. The synthetic core
CC promoter, optionally containing additional upstream elements are used to
CC increase expression, provides non-tissue specific, constitutive
CC transcription of heterologous genes in any sort of plant, especially the
CC gene for oxalate oxidase for increasing plant resistance to pathogens.
CC The upstream activating elements can be used to increase transcription
CC from any promoter. A combination of the synthetic core promoter with
CC synthetic upstream elements can induce expression 10 times greater than

CC that provided by the maize Ubi-1 promoter. The present sequence
CC represents a maize Ubi1 upstream activating region (UAR)
XX
SQL Sequence 813 BP; 241 A; 166 C; 142 G; 264 T; 0 U; 0 Other;

Query Match 100.0%; Score 25; DB 2; Length 813;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTGAGCCCTCTCGAGATTCCGCT 25
DB 654 CTGAGCCCTCTCGAGATTCCGCT 678

RESULT 3
ACC78118
ID ACC78118 standard; DNA; 813 BP.
XX
AC ACC78118;
XX
DT 18-AUG-2003 (first entry)
XX

DE Maize ubiquitin-1 promoter upstream activating region (UAR).
XX
XX Nematode; SCP1; UCP3; SUP; nematocidal; gene therapy; promoter;
XX ubiquitin-1; upstream activating region; UAR; maize; ds.

OS Zea mays.
XX
XX WO2003033651-A2.
XX
XX 24-APR-2003.
XX

PF 08-OCT-2002; 2002WO-US032059.
XX
XX 16-OCT-2001; 2001US-0329667P.
XX

PA (PION-) PIONEER HI-BRED INT INC.
XX

PI Hu X, Lu G, Ruff RL, Schuh W;
XX
XX WPI; 2003-393512/37.
XX

PT New nucleic acid molecule comprising a heterologous nematode-resistance
PT sequence operatively linked to a promoter capable of driving
PT transcripts of the sequence, useful for promoting nematode-resistance
PT in plants.
XX
XX Claim 16; Page 55; 65pp; English.

XX The invention relates to a nucleic acid molecule comprising a
CC heterologous nematode-resistance sequence operably linked to a promoter
CC capable of driving transcripts of the heterologous nematode-resistance
CC sequence in a plant cell, where the promoters are selected from synthetic
CC SCP1, UCP3 and SUP promoters. The nucleic acid molecules and methods are
CC useful for promoting nematode-resistance in plants by inducing expression
CC of a heterologous nematode-resistance sequence, altering expression of a
CC heterologous nucleotide sequence, or creating or enhancing disease
CC resistance in a plant. The present sequence represents the upstream
CC activating region of (UAR) of the ubiquitin-1 promoter from maize
XX

SQL Sequence 813 BP; 241 A; 166 C; 142 G; 264 T; 0 U; 0 Other;

Query Match 100.0%; Score 25; DB 8; Length 813;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTGAGCCCTCTCGAGATTCCGCT 25
DB 654 CTGAGCCCTCTCGAGATTCCGCT 678

RESULT 4

AA211621
 ID AA211621 standard; DNA; 994 BP.
 XX
 AC AA211621;
 XX
 DT 16-NOV-1999 (first entry)
 XX
 DE Ubi1 upstream activating region (UAR) linked with a promoter fragment.
 XX
 KM Plant promoter; TATA motif; transcription start site; upstream element;
 KM gene expression; oxalate oxidase; plant resistance; pathogen; maize;
 KM Ubi-1 promoter; upstream activating region; ss.
 XX
 OS Synthetic.
 OS Zea sp.
 XX
 PN WO9943838-A1.
 XX
 PD 02-SEP-1999.
 XX
 PF 23-FEB-1999; 99WO-US003863.
 XX
 PR 24-FEB-1998; 98US-00028819.
 XX
 PA (PION-) PIONEER HI-BRED INT INC.
 XX
 PI Bowen BA, Bruce WB, Lu G, Sims LE, Tagliani LA;
 XX
 DR WPI; 1999-540601/45.
 XX
 PT New synthetic promoter functional in plants to provide non-tissue
 PT specific, constitutive expression, particularly of oxalate oxidase for
 PT increased resistance to pathogens.
 XX
 PS Disclosure; Page 52-53; 61pp; English.
 XX
 CC The invention provides a new synthetic plant promoter that comprises a
 CC TATA motif; a transcription start site (TSS) and a region between TATA
 CC and TSS containing at least 64 percent GC content. The synthetic core
 CC promoter, optionally containing additional upstream elements are used to
 CC increase expression, provides non-tissue specific, constitutive
 CC transcription of heterologous genes in any sort of plant, especially the
 CC gene for oxalate oxidase for increasing plant resistance to pathogens.
 CC The upstream activating elements can be used to increase transcription
 CC from any promoter. A combination of the synthetic core promoter with
 CC synthetic upstream elements can induce expression 10 times greater than
 CC that provided by the maize Ubi-1 promoter. The present sequence
 CC represents a copy of Ubi1 upstream activating region (UAR) operably
 CC linked with the core promoter
 XX
 SQ Sequence 994 BP; 288 A; 209 C; 181 G; 316 T; 0 U; 0 Other;
 Query Match 100.0%; Score 25; DB 2; Length 994;
 Best Local Similarity 100.0%; Pred. No. 0.16;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTGGAACCCCTCTCGAGAGTTCCGCT 25
 DB 654 CTGGAACCCCTCTCGAGAGTTCCGCT 678
 RESULT 5
 ACC78116
 ID ACC78116 standard; DNA; 1413 BP.
 XX
 AC ACC78116;
 XX
 DT 18-AUG-2003 (first entry)
 XX
 DE Nematode-regulated synthetic promoter SUP.
 XX
 KM Nematode; SCP1, UCP3; SUP; nematocidal; gene therapy; promoter;
 KM 35S promoter; ubiquitin-1; Ryn7-Syn II core promoter; ss.

XX
 OS Synthetic.
 OS Cauliflower mosaic virus.
 OS Zea mays.
 XX
 PN WO2003033651-A2.
 XX
 PD 24-APR-2003.
 XX
 PF 08-OCT-2002; 2002MO-US032059.
 XX
 PR 16-OCT-2001; 2001US-0329667P.
 XX
 PA (PION-) PIONEER HI-BRED INT INC.
 XX
 PI Hu X, Lu G, Ruff RL, Schuh W;
 XX
 DR WPI; 2003-393512/37.
 XX
 PT New nucleic acid molecule comprising a heterologous nematode-resistance
 PT sequence operatively linked to a promoter capable of driving
 PT transcriptions of the sequence, useful for promoting nematode-resistance
 PT in plants.
 XX
 PS Claim 1; Page 54-55; 65pp; English.
 XX
 CC The invention relates to a nucleic acid molecule comprising a
 CC heterologous nematode-resistance sequence operably linked to a promoter
 CC capable of driving transcriptions of the heterologous nematode-resistance
 CC sequence in a plant cell, where the promoters are selected from synthetic
 CC SCP1, UCP3 and SUP promoters. The nucleic acid molecules and methods are
 CC useful for promoting nematode-resistance in plants by inducing expression
 CC of a heterologous nematode-resistance sequence, altering expression of a
 CC heterologous nucleotide sequence, or creating or enhancing disease
 CC resistance in a plant. The present sequence represents a nematode-
 CC regulated promoter SUP1, a synthetic hybrid promoter consisting of one
 CC copy of the upstream activating region of (UAR) of the 35S CamV promoter
 CC from Cauliflower mosaic virus operably linked to the Ryn7 element which
 CC is operably linked to one copy of the UAR of ubiquitin-1 promoter from Z.
 CC mays, which is operably linked to the Ryn7-Syn II core promoter
 XX
 SQ Sequence 1413 BP; 419 A; 309 C; 280 G; 405 T; 0 U; 0 Other;
 Query Match 100.0%; Score 25; DB 8; Length 1413;
 Best Local Similarity 100.0%; Pred. No. 0.16;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTGGAACCCCTCTCGAGAGTTCCGCT 25
 DB 1086 CTGGAACCCCTCTCGAGAGTTCCGCT 1110
 RESULT 6
 AA211622
 ID AA211622 standard; DNA; 1807 BP.
 XX
 AC AA211622;
 XX
 DT 16-NOV-1999 (first entry)
 XX
 DE UCP2 promoter sequence.
 XX
 KM Plant promoter; TATA motif; transcription start site; upstream element;
 KM gene expression; oxalate oxidase; plant resistance; pathogen; maize;
 KM Ubi-1 promoter; upstream activating region; ss.
 XX
 OS Synthetic.
 OS Zea sp.
 XX
 PN WO9943838-A1.
 XX
 PD 02-SEP-1999.
 XX

PF 23-FEB-1999; 99WO-US003863.
XX
XX 24-FEB-1998; 98US-00026819.
XX
XX (PION-) PIONEER HI-BRED INT INC.
XX
XX Bowen BA, Bruce WB, Lu G, Sims LE, Tagliani LA;
XX WPI; 1999-540601/45.
XX
XX
XX New synthetic promoter functional in plants to provide non-tissue
PT specific, constitutive expression, particularly of oxalate oxidase for
PT increased resistance to pathogens.
XX
XX
XX Disclosure; Page 53-54; 61pp; English.
XX
XX The invention provides a new synthetic plant promoter that comprises a
CC TATA motif; a transcription start site (TSS) and a region between TATA
CC and TSS containing at least 64 percent GC content. The synthetic core
CC promoter, optionally containing additional upstream elements are used to
CC increase expression, provides non-tissue specific, constitutive
CC transcription of heterologous genes in any sort of plant, especially the
CC gene for oxalate oxidase for increasing plant resistance to pathogens.
CC The upstream activating elements can be used to increase transcription
CC from any promoter. A combination of the synthetic core promoter with
CC synthetic upstream elements can induce expression 10 times greater than
CC that provided by the maize Ubi-1 promoter. The present sequence
CC represents a UCP2 promoter sequence (2 copies of Ubi1 upstream activating
CC region (UAR) operably linked with the core promoter)
XX
XX Sequence 1807 BP; 530 A; 375 C; 323 G; 579 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 25; DB 2; Length 1807;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTGAGCCCTCTCGAGAGTTCCGCT 25
DB 654 CTGAGCCCTCTCGAGAGTTCCGCT 678
RESULT 7
AD8C4542
ID AD8C4542 standard; DNA; 2005 BP.
XX
XX AC AD8C4542;
XX
XX 01-JAN-2004 (first entry)
XX
XX Modified maize ubiquitin promoter encoding sequence.
XX
XX expression cassette; ds.
XX
XX Zea mays.
XX
XX WO2003027257-A2.
XX
XX 03-APR-2003.
XX
XX 27-SEP-2002; 2002WO-US030895.
XX
XX 27-SEP-2001; 2001US-0325607P.
XX
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
XX Levin JZ, Dietrich R, Budziszewski GJ;
XX WPI; 2003-354651/33.
XX
XX New isolated nucleic acid molecule encoding a polypeptide with a 3'-5'
PT exonuclease domain, useful in molecular biology and transformation, in
PT particular for reproducing and predictably manipulating gene expression
PT in a plant cell.

XX
XX Disclosure; Page 105-106; 108pp; English.
XX
XX
XX The present invention relates to an invention that results in the
CC expression of the endogenous nucleotide sequence in plant cell being
CC increased. In this method the plant cell comprises a second expression
CC cassette comprising a nucleic acid molecule, where the expression of the
CC nucleic acid molecule of interest in the plant cell is decreased as
CC compared to expression of the nucleic acid molecule of interest in a
CC plant cell lacking the first expression cassette. The methods and
CC compositions of the present invention are useful in molecular biology and
CC transformation, in particular for reproducing and predictably
CC manipulating gene expression in a plant cell. The present sequence is a
CC nucleotide sequence of modified maize ubiquitin promoter.
XX
XX
XX Sequence 2005 BP; 474 A; 426 C; 412 G; 693 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 25; DB 10; Length 2005;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTGAGCCCTCTCGAGAGTTCCGCT 25
DB 690 CTGAGCCCTCTCGAGAGTTCCGCT 714
RESULT 8
ACF06306
ID ACF06306 standard; DNA; 2100 BP.
XX
XX AC ACF06306;
XX
XX 07-OCT-2003 (first entry)
XX
XX 5' genome+intact vip3a insert sequence SEQ ID NO:6.
XX
XX Maize; corn; VIP1034; vip3a; pat; vegetative insecticidal protein; VIP;
XX transgenic genotype; insect resistance; herbicide tolerance; corn event;
XX plant; gene; ds.
XX
XX Zea mays.
XX
XX WO2003052073-A2.
XX
XX 26-JUN-2003.
XX
XX 16-DEC-2002; 2002WO-US040099.
XX
XX 17-DEC-2001; 2001US-0341456P.
XX
XX 08-JAN-2002; 2002US-0346660P.
XX
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
XX Steiner H, Dawson J, Dunder E, Meghji M;
XX WPI; 2003-533014/50.
XX
XX New transgenic corn event, VIP1034, useful for conferring a
PT characteristic consisting of insect resistance and herbicide tolerance.
PT
PT Claim 7; Page 56-57; 73pp; English.
XX
XX The present invention describes an isolated nucleic acid molecule (1)
CC comprising a sequence having at least one junction sequence of corn event
CC VIP1034 comprising a sequence having 20 bp, or its complement. VIP1034 is
CC a vegetative insecticidal protein (VIP), and comprises a novel transgenic
CC genotype that comprises a vip3a gene and a pat gene, which confers insect
CC resistance and herbicide tolerance to the transgenic corn event. Also
CC described: (1) an amplicon comprising (i); (2) a polynucleotide primer
CC for detecting corn event VIP1034 in a sample comprising at least 15
CC contiguous nucleotides from position 1-716 of the 2100-bp sequence given
CC in ACF06306 or its complement; (3) detecting the presence of a DNA
CC corresponding to the VIP1034 event in the sample; (4) a kit for detecting

CC the presence of VIP1034 nucleic acids in a sample; (5) detecting corn
CC event VIP1034 in a biological sample; (6) a seed of any corn inbred
CC comprising the transgenic genotype of the corn event VIP1034, where the
CC transgenic genotype comprises an intact copy and a fragmented copy of a
CC first expression cassette and an intact copy of a second expression
CC cassette; (7) a corn plant or its pollen or ovule comprising the
CC transgenic genotype of the corn event VIP1034; (8) a hybrid corn seed
CC produced by crossing the corn plant with an inbred corn plant having a
CC different genotype; (9) a hybrid corn plant produced by growing the
CC hybrid corn seed; (10) producing corn seed; and (11) producing hybrid
CC corn seeds. The nucleic acid is useful for conferring a characteristic
CC consisting of insect resistance and herbicide tolerance. The present
CC sequence represents a corn 5' genome-intact vip3A insert sequence, which
CC is used in the exemplification of the present invention

SQ Sequence 2100 BP; 555 A; 479 C; 397 G; 668 T; 0 U; 1 Other;

Query Match 100.0%; Score 25; DB 9; Length 2100;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGACCCCTCTCGAGAGTTCGCT 25
DB 1401 CTGACCCCTCTCGAGAGTTCGCT 1425

RESULT 9
ACC78115
ID ACC78115 standard; DNA; 2609 BP.
XX
AC ACC78115;
XX
DT 18-AUG-2003 (first entry)
XX
DE Nematode-regulated synthetic promoter UCP3.
XX
KW Nematode; SCP1; UCP3; SUP; nematocidal; gene therapy; promoter;
KW ubiquitin-1; Reyn7-Syn II core promoter; ss.
XX
OS Synthetic.
OS Zea mays.
XX
PN WO2003033651-A2.
XX
PD 24-APR-2003.
XX
PF 08-OCT-2002; 2002WO-US032059.
XX
PR 16-OCT-2001; 2001US-0329667P.
XX
PA (PION-) PIONEER HI-BRED INT' INC.
XX
PI Hu X, Lu G, Ruff RL, Schuh W;
XX
DR WPI; 2003-393512/37.
XX
PT New nucleic acid molecule comprising a heterologous nematode-resistance
PT sequence operatively linked to a promoter capable of driving
PT transcriptions of the sequence, useful for promoting nematode-resistance
PT in plants.
XX
PS Claim 1; Page 53-54; 65pp; English.
XX
XX The invention relates to a nucleic acid molecule comprising a
CC heterologous nematode-resistance sequence operably linked to a promoter
CC capable of driving transcriptions of the heterologous nematode-resistance
CC sequence in a plant cell, where the promoters are selected from synthetic
CC SCP1, UCP3 and SUP promoters. The nucleic acid molecules and methods are
CC useful for promoting nematode-resistance in plants by inducing expression
CC of a heterologous nematode-resistance sequence, altering expression of a
CC heterologous nucleotide sequence, or creating or enhancing disease
CC resistance in a plant. The present sequence represents a nematode-
CC regulated promoter UCP3, a synthetic hybrid promoter consisting of three

CC copies (in forward orientation) of the upstream activating region (UAR)
CC of the ubiquitin-1 promoter from Z. mays operably linked to the Reyn7-Syn
CC II core promoter
XX
SQ Sequence 2609 BP; 769 A; 542 C; 464 G; 834 T; 0 U; 0 Other;

Query Match 100.0%; Score 25; DB 8; Length 2609;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGACCCCTCTCGAGAGTTCGCT 25
DB 654 CTGACCCCTCTCGAGAGTTCGCT 678

RESULT 10
AAZ11623
ID AAZ11623 standard; DNA; 2620 BP.
XX
AC AAZ11623;
XX
DT 16-NOV-1999 (first entry)
XX
DE Ubi1 upstream activating region (UAR) linked with a promoter fragment.
XX
KW Plant promoter; TATA motif; transcription start site; upstream element;
KW gene expression; oxalate oxidase; plant resistance; pathogen; maize;
KW Ubi-1 promoter; upstream activating region; ss.
XX
OS Synthetic.
OS Zea sp.
XX
PN WO9943838-A1.
XX
PD 02-SEP-1999.
XX
PF 23-FEB-1999; 99WO-US003863.
XX
PR 24-FEB-1998; 98US-00028819.
XX
PA (PION-) PIONEER HI-BRED INT' INC.
XX
PI Bowen BA, Bruce WB, Lu G, Sims LE, Tagliani LA;
XX
DR WPI; 1999-540601/45.
XX
PT New synthetic promoter functional in plants to provide non-tissue
PT specific, constitutive expression, particularly of oxalate oxidase for
PT increased resistance to pathogens.
XX
PS Disclosure; Page 54-56; 61pp; English.
XX
XX The invention provides a new synthetic plant promoter that comprises a
CC TATA motif; a transcription start site (TSS) and a region between TATA
CC and TSS containing at least 64 percent GC content. The synthetic core
CC promoter, optionally containing additional upstream elements are used to
CC increase expression, provides non-tissue specific, constitutive
CC transcription of heterologous genes in any sort of plant, especially the
CC gene for oxalate oxidase for increasing plant resistance to pathogens.
CC The upstream activating elements can be used to increase transcription
CC from any promoter. A combination of the synthetic core promoter with
CC synthetic upstream elements can induce expression 10 times greater than
CC that provided by the maize Ubi-1 promoter. The present sequence
CC represents a copy of Ubi1 upstream activating region (UAR) operably
CC linked with the core promoter
XX
SQ Sequence 2620 BP; 772 A; 541 C; 465 G; 842 T; 0 U; 0 Other;

Query Match 100.0%; Score 25; DB 2; Length 2620;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGACCCCTCTCGAGAGTTCGCT 25

Db 654 CTGACCCCTCTCGAGATTCCGCT 678

RESULT 11

AAH19412

AAH19412 standard; DNA; 2807 BP.

AC AAH19412;

30-JUL-2001 (first entry)

SUC-1 promoter element.

Promoter; SUC-1; transgenic plant; CamV 35S promoter; maize; polyubiquitin promoter; UBI promoter; sugarcane gene expression; monocotyledonous plant gene expression; ds.

OS Synthetic.

Key Location/Qualifiers

5'UTR 1715..1797

/tag= a

/note= "Maize UBI 5' UTR"

intron 1798..2807

/tag= c

WO200132897-A2.

10-MAY-2001.

06-NOV-2000; 2000WO-IB001601.

05-NOV-1999; 99ZA-00006954.

11-MAY-2000; 2000ZA-00002313.

(SASU-) SOUTH AFRICAN SUGAR ASSOC.

(UVST-) UNIV STELLENBOSCH INST PLANT BIOTECHNOLO.

Groenewald J, Botha FC;

WPI; 2001-316449/33.

A new promoter construct designated SUC-1 contains promoter elements from CamV 35S and maize UBI and is useful to confer stable high level constitutive expression in transgenic plants.

Claim 1; Fig 1; 19pp; English.

The present sequence is tandem promoter construct, SUC-1. This promoter confers stable, high level, constitutive expression in transgenic plants. The promoter contains both the cauliflower mosaic virus (CaMV) 35S promoter and the maize polyubiquitin (UBI) promoter. The promoter is particularly useful for gene expression in monocotyledonous plants, such as sugarcane. The combination of the two promoters gives higher expression levels in monocotyledonous plants than the CamV 35S or UBI promoter when used alone

Sequence 2807 BP; 751 A; 607 C; 582 G; 867 T; 0 U; 0 Other;

Query Match 100.0%; Score 25; DB 4; Length 2807;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CTGACCCCTCTCGAGATTCCGCT 25

1501 CTGACCCCTCTCGAGATTCCGCT 1525

RESULT 12
ACF06318/c
ID ACF06318 standard; DNA; 2848 BP.
XX

AC ACF06318;

07-OCT-2003 (first entry)

Pat insect nucleotide sequence SEQ ID NO:18.

Maize; corn; VIP1034; vip3a; pat; vegetative insecticidal protein; VIP; transgenic genotype; insect resistance; herbicide tolerance; corn event; plant; gene; ds.

Zea mays.

WO2003052073-A2.

26-JUN-2003.

16-DEC-2002; 2002WO-US040099.

17-DEC-2001; 2001US-0341456P.

08-JAN-2002; 2002US-0346660P.

(SYGN) SYNGENTA PARTICIPATIONS AG.

Steiner H, Dawson J, Dunder E, Meghji M;

WPI; 2003-533014/50.

New transgenic corn event, VIP1034, useful for conferring a characteristic consisting of insect resistance and herbicide tolerance.
Claim 22; Page 68-69; 73pp; English.

The present invention describes an isolated nucleic acid molecule (I) comprising a sequence having at least one junction sequence of corn event VIP1034 comprising a sequence having 20 bp, or its complement. VIP1034 is a vegetative insecticidal protein (VIP), and comprises a novel transgenic genotype that comprises a vip3a gene and a pat gene, which confers insect resistance and herbicide tolerance to the transgenic corn event. Also described: (1) an amplicon comprising (1); (2) a polynucleotide primer for detecting corn event VIP1034 in a sample comprising at least 15 contiguous nucleotides from position 1-716 of the 2100-bp sequence given in ACF06318 or its complement; (3) detecting the presence of a DNA corresponding to the VIP1034 event in the sample; (4) a kit for detecting the presence of VIP1034 nucleic acids in a sample; (5) detecting corn event VIP1034 in a biological sample; (6) a seed of any corn inbred comprising the transgenic genotype of the corn event VIP1034, where the transgenic genotype comprises an intact copy and a fragmented copy of a cassette; (7) a corn plant or its pollen or ovule comprising the transgenic genotype of the corn event VIP1034; (8) a hybrid corn seed produced by crossing the corn plant with an inbred corn plant having a different genotype; (9) a hybrid corn plant produced by growing the hybrid corn seed; (10) producing corn seed, and (11) producing hybrid corn seeds. The nucleic acid is useful for conferring a characteristic consisting of insect resistance and herbicide tolerance. The present sequence represents a corn pat insect nucleotide sequence, which is used in the exemplification of the present invention

Sequence 2848 BP; 937 A; 632 C; 580 G; 699 T; 0 U; 0 Other;

Query Match 100.0%; Score 25; DB 9; Length 2848;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CTGACCCCTCTCGAGATTCCGCT 25

2159 CTGACCCCTCTCGAGATTCCGCT 2135

RESULT 13
AAD11579
ID AAD11579 standard; DNA; 3069 BP.
XX

CC within a target gene and allowing the ZFP binding to the target
 CC nucleotide sequence, where the expression of the target gene in the plant
 CC cells is modulated. The ZFP and fusions of the ZFP proteins are useful
 CC for modulating or regulating gene expression and metabolic pathways in
 CC plants. The ZFP, fusion proteins and methods are useful in plant and
 CC agricultural technology. The method is useful particularly for treating a
 CC disorder in the plant cells, where the disorder is associated with
 CC abnormal expression of the target gene. The present DNA sequence is
 CC pND3008 activation construct. This sequence is used to construct maize
 CC ubiquitin promoter Zmubi effector construct which is further used to
 CC determine the function of ZFP-effector fusion protein on luciferase
 CC reporter gene in maize cells

SQ Sequence 3121 BP; 755 A; 713 C; 684 G; 969 T; 0 U; 0 Other;
 Query Match 100.0%; Score 25; DB 4; Length 3121;
 Best Local Similarity 100.0%; Pred. No. 0.17;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGAGCCCTCTCGAGAGTTCGCT 25
 DB 677 CTGAGCCCTCTCGAGAGTTCGCT 701

RESULT 15

AAZ11624
 ID AAZ11624 standard; DNA; 3433 BP.

AAZ11624;

16-NOV-1999 (first entry)

UCP4 promoter sequence.

XX Plant promoter; TATA motif; transcription start site; upstream element;
 KW gene expression; oxalate oxidase; plant resistance; pathogen; maize;
 KW Ubi-1 promoter; upstream activating region; ss.

XX Synthetic.

OS Zea sp.

PN MO9943838-A1.

PD 02-SEP-1999.

PF 23-FEB-1999; 99WO-US003863.

PR 24-FEB-1998; 98US-00028819.

PA (PION-) PIONEER HI-BRED INT INC.

PI Bowen BA, Bruce WB, Lu G, Sims LE, Tagliani LA;

DR WPI; 1999-540601/45.

PT New synthetic promoter functional in plants to provide non-tissue
 PT specific, constitutive expression, particularly of oxalate oxidase for
 PT increased resistance to pathogens.

PS Disclosure; Page 56-58; 61pp; English.

CC The invention provides a new synthetic plant promoter that comprises a
 CC TATA motif; a transcription start site (TSS) and a region between TATA
 CC and TSS containing at least 64 percent GC content. The synthetic core
 CC promoter, optionally containing additional upstream elements are used to
 CC increase expression, provides non-tissue specific, constitutive
 CC transcription of heterologous genes in any sort of plant, especially the
 CC gene for oxalate oxidase for increasing plant resistance to pathogens.
 CC The upstream activating elements can be used to increase transcription
 CC from any promoter. A combination of the synthetic core promoter with
 CC synthetic upstream elements can induce expression 10 times greater than
 CC that provided by the maize Ubi-1 promoter. The present sequence
 CC represents a UCP4 promoter sequence (4 copies of Ubi1 upstream activating

CC region (UAR) operably linked with the core promoter)
 XX SQ Sequence 3433 BP; 1014 A; 707 C; 607 G; 1105 T; 0 U; 0 Other;

Query Match 100.0%; Score 25; DB 2; Length 3433;
 Best Local Similarity 100.0%; Pred. No. 0.17;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGAGCCCTCTCGAGAGTTCGCT 25
 DB 654 CTGAGCCCTCTCGAGAGTTCGCT 678

Search completed: February 28, 2005, 23:51:37
 Job time : 223.182 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: February 28, 2005, 23:35:01 ; Search time 1734.42 Seconds
(without alignment)
548.661 Million cell updates/sec

Title: US-10-086-062-1

Perfect score: 25
Sequence: 1 ctgacccctctcagagatccgc 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_esc1:*
2: gb_esc2:*
3: gb_hic:*
4: gb_esc3:*
5: gb_esc4:*
6: gb_esc5:*
7: gb_esc6:*
8: gb_gsa1:*
9: gb_gsa2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	229	8	BH638797 1008024H1
2	25	100.0	307	9	CG805015 118056F0
3	25	100.0	338	9	CG805036 118056F0
4	25	100.0	534	8	BH638999 1008026F0
5	25	100.0	617	9	CG709678 119014D0
6	25	100.0	620	9	CG805054 118056H0
7	25	100.0	716	9	CG472745 GQ0081.BR
8	25	100.0	765	7	CG472714 GQ0081.BR
9	25	100.0	769	7	CG472736 GQ0081.BR
10	25	100.0	774	7	CG472678 GQ0081.BR
11	25	100.0	805	7	CG472728 GQ0081.BR
12	25	100.0	817	7	CG472729 GQ0081.BR
13	25	100.0	819	7	CG472694 GQ0081.BR
14	25	100.0	820	7	CG472674 GQ0081.BR
15	25	100.0	825	7	CG472692 GQ0081.BR
16	25	100.0	828	7	CG472722 GQ0081.BR
17	25	100.0	828	7	CG472731 GQ0081.BR
18	25	100.0	833	7	CG472737 GQ0081.BR
19	25	100.0	837	7	CG472693 GQ0081.BR
20	25	100.0	841	7	CG472673 GQ0081.BR
21	25	100.0	851	7	CG472751 GQ0081.BR
22	25	100.0	852	7	CG472735 GQ0081.BR
23	25	100.0	852	7	CG472741 GQ0081.BR
24	25	100.0	856	7	CG472725 GQ0081.BR

25	25	100.0	859	7	CG472744 GQ0081.BR
26	25	100.0	876	7	CG472681 GQ0081.BR
27	25	100.0	879	7	CG472711 GQ0081.BR
28	25	100.0	940	9	CG670569 CGFWM88TV
29	25	100.0	965	9	CG342774 CGYB89TV
30	23	92.0	794	7	CG472764 GQ0081.BR
31	23	92.0	853	7	CG472754 GQ0081.BR
32	21	84.0	720	8	CK438248 GQ0081.BR
33	20.2	80.8	100	8	BH219170 1006085A0
34	20.2	80.8	415	4	BM661080 952044B06
35	20.2	80.8	553	4	BM736979 952051F11
36	20.2	80.8	636	9	BI679714 949078B10
37	20.2	80.8	728	4	CG212795 CGXCY79TV
38	20.2	80.8	896	9	CG212788 CGXCY79TV
39	19.8	79.2	300	6	C55594 C55594 Yuj1
40	19.2	76.8	669	9	CL173701 104.377.1
41	19.2	76.8	897	5	BU838185 AGENCOURT
42	19.2	76.8	908	2	BE797311 601587675
43	19.2	76.8	1283	8	AZ130054 OSJNB010
44	18.8	75.2	323	4	BI029471 IL5-MT026
45	18.8	75.2	472	2	BE142500 RC6-HT014

ALIGNMENTS

RESULT 1
BH638797 1008024H1.2EL_x1 1008 - Rescemu Grid I Zea mays genomic, genomic survey sequence.
BH638797 1008024H1.2EL_x1 1008 - Rescemu Grid I Zea mays genomic, genomic survey sequence.
BH638797.1 GI:18663263

ACCESSION
LOCUS
DEFINITION
VERSION
KEYWORDS
SOURCE
ORGANISM
Zea mays
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 229)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Walbot, V.
Maize genomic sequences found using engineered Rescemu transposon
Unpublished (2001)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221

Email: walbot@stanford.edu
Possible ligation site of ends cut by 2 different endonucleases.
Reverse complemented post-ligation sequence from source sequence.
Plate: 1008024 row: 10
Classes: transposon-tagged.
Location/Qualifiers
1..229

FEATURES
source

/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="mixed background W23/A188/B73"
/db_xref="taxon:4577"
/tissue_type="leaf"
/dev_stage="adult"
/lab_host="DH10B"
/clone_11b="1008 - Rescemu Grid I"
/note="Organ: leaf; Vector: Rescemu (engineered from pBluescript backbone); Site_1: BamHI; Site_2: BglII; Rescemu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on Rescemu, go to the web site www.zmdb.laastate.edu and follow the links for 'Rescemu'. Grid I was grown at Berkeley in 2001. DNA was extracted from leaf punches, double digested using BamHI

and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."

ORIGIN

Query Match 100.0%; Score 25; DB 8; Length 229;
Best Local Similarity 100.0%; Pred. No. 0.64;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTGAGACCCCTCTCGAGAGTTCCGCT 25
DB 25 CTGAGACCCCTCTCGAGAGTTCCGCT 49

RESULT 2
LOCUS CG805015 307 bp DNA linear GSS 10-NOV-2003
DEFINITION 118056F09.2EL_x1 1118 - RescuenMu Grid S Zea mays genomic, genomic survey sequence.
ACCESSION CG805015 GI:38241016
VERSION CG805015.1
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 307)
Walbot, V.
Maize genomic sequences found using engineered RescuenMu transposon
REFERENCE Unpublished (2001)
AUTHORS Contact: Walbot V
TITLE Department of Biological Sciences
JOURNAL Stanford University
COMMENT 855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Possible ligation site of ends cut by 2 different endonucleases.
Reverse complemented post-ligation sequence from source sequence.
Plate: 1118056 row: 40
Class: transposon-tagged.
Location/Qualifiers
1..307
/organism="Zea mays"
/mol_type="genomic DNA"
/culturvar="mixed background W23/A188/B73"
/db_xref="taxon:4577"
/tissue_type="leaf"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="1118 - RescuenMu Grid S"
/note="Organ: leaf; Vector: RescuenMu (engineered from pBluescript backbone); Site 1: BamHI; Site 2: BglII; RescuenMu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on RescuenMu, go to the web site 'www.zmdb.iastate.edu' and follow the links for 'RescuenMu'. Grid S was grown at San Diego in 2002. DNA was extracted from leaf strips, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."

FEATURES

source

ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 307;
Best Local Similarity 100.0%; Pred. No. 0.65;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTGAGACCCCTCTCGAGAGTTCCGCT 25
DB 80 CTGAGACCCCTCTCGAGAGTTCCGCT 104

RESULT 3
LOCUS CG805036 338 bp DNA linear GSS 10-NOV-2003
DEFINITION 118056F03.2EL_x1 1118 - RescuenMu Grid S Zea mays genomic, genomic survey sequence.
ACCESSION CG805036
VERSION CG805036.1 GI:38241053
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 338)
Walbot, V.
Maize genomic sequences found using engineered RescuenMu transposon
REFERENCE Unpublished (2001)
AUTHORS Contact: Walbot V
TITLE Department of Biological Sciences
JOURNAL Stanford University
COMMENT 855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Possible ligation site of ends cut by 2 different endonucleases.
Reverse complemented post-ligation sequence from source sequence.
Plate: 1118056 row: 40
Class: transposon-tagged.
Location/Qualifiers
1..338
/organism="Zea mays"
/mol_type="genomic DNA"
/culturvar="mixed background W23/A188/B73"
/db_xref="taxon:4577"
/tissue_type="leaf"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="1118 - RescuenMu Grid S"
/note="Organ: leaf; Vector: RescuenMu (engineered from pBluescript backbone); Site 1: BamHI; Site 2: BglII; RescuenMu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on RescuenMu, go to the web site 'www.zmdb.iastate.edu' and follow the links for 'RescuenMu'. Grid S was grown at San Diego in 2002. DNA was extracted from leaf strips, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."

FEATURES

source

ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 338;
Best Local Similarity 100.0%; Pred. No. 0.65;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTGAGACCCCTCTCGAGAGTTCCGCT 25
DB 78 CTGAGACCCCTCTCGAGAGTTCCGCT 102

RESULT 4
LOCUS BH638999 534 bp DNA linear GSS 14-FEB-2002
DEFINITION 1008026F08.1EL_x1 1008 - RescuenMu Grid I Zea mays genomic, genomic survey sequence.
ACCESSION BH638999 GI:18663697
VERSION BH638999.1
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
TITLE Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
AUTHORS 1 (bases 1 to 534)
JOURNAL Maize genomic sequences found using engineered Rescemu transposon
COMMENT Unpublished (2001)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Very probable ligation site of ends cut by single endonuclease.
Reverse complemented post-ligation sequence from source sequence.
Plate: 108026 row: 10
Class: transposon-tagged.
Location/Qualifiers

FEATURES
source
1..534
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="mixed background W23/A188/B73"
/db_xref="taxon:4577"
/tissue_type="leaf"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="1008 - Rescemu Grid I"
/note="Organ: leaf; Vector: Rescemu (engineered from
pBluescript backbone); Site 1: BamHI; Site 2: BglII;
Rescemu is a 4.9 kb, modified maize Mu transposon
designed to allow plasmid rescue from total genomic DNA.
Mu elements insert preferentially into transcription
units. For more information on Rescemu, go to the web
site www.zmdb.iastate.edu and follow the links for
'Rescemu.' Grid I was grown at Berkeley in 2001. DNA was
extracted from leaf punches, double digested using BamHI
and BglII, and ligated to form circular plasmids. DH10B
cells were transformed and then screened on LB plates with
ampicillin."

ORIGIN
Query Match 100.0%; Score 25; DB 8; Length 534;
Best Local Similarity 100.0%; Pred. No. 0.67; 0; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTGACCCCTCTCGAGATTCGCT 25
|||||
Db 246 CTGACCCCTCTCGAGATTCGCT 270

RESULT 5
LOCUS CG709678 617 bp DNA linear GSS 20-OCT-2003
DEFINITION 1119014D02.2EL.X1 1119 - Rescemu Grid AA Zea mays genomic, genomic
survey sequence.
ACCESSION CG709678
VERSION CG709678.1 GI:37735584
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 617)
Email: walbot@stanford.edu
Possible ligation site of ends cut by 2 different endonucleases.
Reverse complemented post-ligation sequence from source sequence.
Plate: 1118056 row: 40
Class: transposon-tagged.
Location/Qualifiers
1..620
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="mixed background W23/A188/B73"
/db_xref="taxon:4577"

Fax: 650 725 8221
Email: walbot@stanford.edu
Possible ligation site of ends cut by 2 different endonucleases.
Reverse complemented post-ligation sequence from source sequence.
Plate: 1119014 row: 29
Class: transposon-tagged.
Location/Qualifiers
1..617
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="mixed background W23/A188/B73/K55"
/db_xref="taxon:4577"
/tissue_type="leaf"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="1119 - Rescemu Grid AA"
/note="Organ: leaf; Vector: Rescemu (engineered from
pBluescript backbone); Site 1: BamHI; Site 2: BglII;
Rescemu is a 4.9 kb, modified maize Mu transposon
designed to allow plasmid rescue from total genomic DNA.
Mu elements insert preferentially into transcription
units. For more information on Rescemu, go to the web
site www.zmdb.iastate.edu and follow the links for
'Rescemu.' Grid AA was grown at UC San Diego in 2002. DNA
was extracted from leaf strips, double digested using
BamHI and BglII, and ligated to form circular plasmids.
DH10B cells were transformed and then screened on LB
plates with ampicillin."

ORIGIN
Query Match 100.0%; Score 25; DB 9; Length 617;
Best Local Similarity 100.0%; Pred. No. 0.68; 0; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTGACCCCTCTCGAGATTCGCT 25
|||||
Db 356 CTGACCCCTCTCGAGATTCGCT 380

RESULT 6
LOCUS CG805054 620 bp DNA linear GSS 10-NOV-2003
DEFINITION 1118056H01.2EL.X1 1118 - Rescemu Grid S Zea mays genomic, genomic
survey sequence.
ACCESSION CG805054
VERSION CG805054.1 GI:38241084
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 620)
Email: walbot@stanford.edu
Possible ligation site of ends cut by 2 different endonucleases.
Reverse complemented post-ligation sequence from source sequence.
Plate: 1118056 row: 40
Class: transposon-tagged.
Location/Qualifiers
1..620
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="mixed background W23/A188/B73"
/db_xref="taxon:4577"

REFERENCE Maize genomic sequences found using engineered Rescemu transposon
AUTHORS Unpublished (2001)
JOURNAL Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Possible ligation site of ends cut by 2 different endonucleases.
Reverse complemented post-ligation sequence from source sequence.
Plate: 1118056 row: 40
Class: transposon-tagged.
Location/Qualifiers
1..620
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="mixed background W23/A188/B73"
/db_xref="taxon:4577"

/tissue_type="leaf"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="1118 - RescueMu Grid 8"
/note="Organ: leaf; Vector: RescueMu (engineered from pBluescript backbone); Site 1: BamHI, Site 2: BglII; RescueMu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on RescueMu, go to the web site 'www.zmblabstate.edu' and follow the links for 'RescueMu'. Grid 8 was grown at San Diego in 2002. DNA was extracted from leaf strips, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."

ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 620;
Best Local Similarity 100.0%; Pred. No. 0.68;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGGACCCCTCTCGAGAGTTCCGCT 25
|||||
Db 358 CTGGACCCCTCTCGAGAGTTCCGCT 382

RESULT 7

CO472745

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

CO472745 716 bp mRNA linear EST 09-JUL-2004
GQ0081.BR.1.D19 GQ008: Cambium, phloem and bark of girdled saplings
CO472745 Picea glauca cDNA clone GQ0081.BR.D19 5', mRNA sequence.
CO472745.1 GI:50141269
EST.
Picea glauca (white spruce)
Picea glauca
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
1 (bases 1 to 716)
Morency,M.-J., Cooke,J., Pavy,N., Parsons,L., Paule,C., Seguin,A.,
Retzell,E., Butterfield,Y., Barber,S., Yang,G., Scott,J.,
Siddiqui,A., Holt,R., Marra,M. and Mackay,J.
Arborea EST sequencing in Picea glauca (white spruce)
Unpublished (2004)
Contact: John Mackay
Centre de Recherche en Biologie Forestiere
Universite Laval
Pavillon Charles-Eugene Marchand, Quebec, Quebec, CANADA G1K 7P4
Fax: 418 656 7493
Email: jmackay@rsb.ulaval.ca
Center for Computational Genomics and Bioinformatics (CCGB),
University of Minnesota, MN id identifier: MN519733 Clone ID:
GQ0081.BR.D19 Clones available through: John Mackay, Ph. D.
jmackey@rsb.ulaval.ca Assistant professor EMAIL:
jmackey@rsb.ulaval.ca Centre de Recherche en Biologie Forestiere
(Forest Biology Research Center) Universite Laval Quebec, Quebec
CANADA G1K 7P4
Plate: 1.BR row: 19 column: D
Seq primer: M13 Reverse Primer.
Location/Qualifiers
1..716
/organism="Picea glauca"
/mol_type="mRNA"
/strain="Pg-653"
/db_xref="taxon:3330"
/clone="GQ0081.BR.D19"
/sex="Hermaphrodite"
/tissue_type="Vascular cambium, secondary phloem and bark
of trees girdled by removing a ring of bark ca. 1 cm wide
from the midpoint of the main stem"
/dev_stage="Vascular cambium, secondary phloem, and bark
from trees harvested 1 and 7 days after girdling

treatment"
/lab_host="E. coli DH10B cells"
/clone_lib="GQ008: Cambium, phloem and bark of girdled
saplings"
/note="Organ: Main stem of 4 year old saplings,
approximately 60 cm tall; Vector: pBluescript II SK (+)
XR; Site 1: EcoRI; Site 2: Xho-I; Tissues were pooled
from above and below the girdle. cDNA was prepared from 5
mg of poly A+ selected RNA and was directionally ligated
into the pBluescript II SK (+) XR vector (Stratagene),
transformed by electroporation into DH10B cells (in
vitrogen) for propagation"

ORIGIN

Query Match 100.0%; Score 25; DB 7; Length 716;
Best Local Similarity 100.0%; Pred. No. 0.68;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGGACCCCTCTCGAGAGTTCCGCT 25
|||||
Db 666 CTGGACCCCTCTCGAGAGTTCCGCT 690

RESULT 8

CO472714

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

CO472714 765 bp mRNA linear EST 09-JUL-2004
GQ0081.BR.1.C16 GQ008: Cambium, phloem and bark of girdled saplings
CO472714 Picea glauca cDNA clone GQ0081.BR.C16 5', mRNA sequence.
CO472714.1 GI:50141206
EST.
Picea glauca (white spruce)
Picea glauca
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
1 (bases 1 to 765)
Morency,M.-J., Cooke,J., Pavy,N., Parsons,L., Paule,C., Seguin,A.,
Retzell,E., Butterfield,Y., Barber,S., Yang,G., Scott,J.,
Siddiqui,A., Holt,R., Marra,M. and Mackay,J.
Arborea EST sequencing in Picea glauca (white spruce)
Unpublished (2004)
Contact: John Mackay
Centre de Recherche en Biologie Forestiere
Universite Laval
Pavillon Charles-Eugene Marchand, Quebec, Quebec, CANADA G1K 7P4
Fax: 418 656 7493
Email: jmackay@rsb.ulaval.ca
Center for Computational Genomics and Bioinformatics (CCGB),
University of Minnesota, MN id identifier: MN5197864 Clone ID:
GQ0081.BR.C16 Clones available through: John Mackay, Ph. D.
jmackey@rsb.ulaval.ca Assistant professor EMAIL:
jmackey@rsb.ulaval.ca Centre de Recherche en Biologie Forestiere
(Forest Biology Research Center) Universite Laval Quebec, Quebec
CANADA G1K 7P4
Plate: 1.BR row: 16 column: C
Seq primer: M13 Reverse Primer.
Location/Qualifiers
1..765
/organism="Picea glauca"
/mol_type="mRNA"
/strain="Pg-653"
/db_xref="taxon:3330"
/clone="GQ0081.BR.C16"
/sex="Hermaphrodite"
/tissue_type="Vascular cambium, secondary phloem and bark
of trees girdled by removing a ring of bark ca. 1 cm wide
from the midpoint of the main stem"
/dev_stage="Vascular cambium, secondary phloem, and bark
from trees harvested 1 and 7 days after girdling
treatment"
/lab_host="E. coli DH10B cells"
/clone_lib="GQ008: Cambium, phloem and bark of girdled
saplings"

/note="Organ: Main stem of 4 year old saplings, approximately 60 cm tall; Vector: pBluescript II SK (+) XR; Site 1: Eco-RI; Site 2: Xho-I; Tissues were pooled from above and below the girdle. cDNA was prepared from 5 mg of poly A+ selected RNA and was directionally ligated into the pBluescript II SK (+) XR vector (Stratagene), transformed by electroporation into DH10B cells (in vitro) for propagation"

ORIGIN

Query Match 100.0%; Score 25; DB 7; Length 765;
Best Local Similarity 100.0%; Pred. No. 0.68;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTGACCCCTCTCGAGAGTTCGGCT 25
Db 671 CTGACCCCTCTCGAGAGTTCGGCT 695

RESULT 9
LOCUS CO472736 769 bp mRNA linear EST 09-JUL-2004
DEFINITION GQ0081.BR.1.G06 GQ008: Cambium, phloem and bark of girdled saplings
Picea glauca cDNA clone GQ0081.BR_G06 5', mRNA sequence.
CO472736
ACCESSION CO472736.1 GI:50141251
VERSION
KEYWORDS
SOURCE EST.
ORGANISM Picea glauca (white spruce)

REFERENCE
AUTHORS Bukurycota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
1 (bases 1 to 769)
Morency, M.-J., Cooke, J., Pavy, N., Parsons, L., Paule, C., Seguin, A., Retzel, E., Butterfield, Y., Barber, S., Yang, G., Stott, J., Sidiqi, A., Holt, R., Maïra, M. and Mackay, J.
Arborea EST sequencing in Picea glauca (white spruce)
Unpublished (2004)
JOURNAL
TITLE Contact: John Mackay
COMMENT Centre de Recherche en Biologie Forestiere
Universite Laval
Pavillon Charles-Eugene Marchand, Quebec, CANADA G1K 7P4
Fax: 418 656 7493
Email: jmackay@rsvs.ulaval.ca
Center for Computational Genomics and Bioinformatics (CCGB),
University of Minnesota, MN Id Identifier: MN5198051 Clone ID:
GQ0081.BR.G06 Clones available through: John Mackay, Ph. D.
Professeur adjoint -Assistant professeur EMAIL:
jmackay@rsvs.ulaval.ca Centre de Recherche en Biologie Forestiere
(Forest Biology Research Center) Universite Laval Quebec, Quebec
CANADA G1K 7P4
Plate: 1.BR row: 06 column: G
Seq primer: M3 Reverse Primer.
Location/Qualifiers

FEATURES

source
1..769
/organism="Picea glauca"
/mol_type="mRNA"
/strain="pg-653"
/db_xref="taxon:3330"
/clone="GQ0081.BR.G06"
/sex="Hermaphrodite"
/tissue_type="Vascular cambium, secondary phloem and bark of trees girdled by removing a ring of bark ca. 1 cm wide from the midpoint of the main stem"
/dev_stage="Vascular cambium, secondary phloem, and bark from trees harvested 1 and 7 days after girdling treatment"
/lab_host="E. coli DH10B cells"
/clone_1lb="GQ008: Cambium, phloem and bark of girdled saplings"
/note="Organ: Main stem of 4 year old saplings, approximately 60 cm tall; Vector: pBluescript II SK (+) XR; Site 1: Eco-RI; Site 2: Xho-I; Tissues were pooled from above and below the girdle. cDNA was prepared from 5

mg of poly A+ selected RNA and was directionally ligated into the pBluescript II SK (+) XR vector (Stratagene), transformed by electroporation into DH10B cells (in vitro) for propagation"

ORIGIN

Query Match 100.0%; Score 25; DB 7; Length 769;
Best Local Similarity 100.0%; Pred. No. 0.68;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTGACCCCTCTCGAGAGTTCGGCT 25
Db 675 CTGACCCCTCTCGAGAGTTCGGCT 699

RESULT 10
LOCUS CO472678 774 bp mRNA linear EST 09-JUL-2004
DEFINITION GQ0081.BR.1.D17 GQ008: Cambium, phloem and bark of girdled saplings
Picea glauca cDNA clone GQ0081.BR_D17 5', mRNA sequence.
CO472678
ACCESSION CO472678.1 GI:50141128
VERSION
KEYWORDS
SOURCE EST.
ORGANISM Picea glauca (white spruce)

REFERENCE
AUTHORS Bukurycota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
1 (bases 1 to 774)
Morency, M.-J., Cooke, J., Pavy, N., Parsons, L., Paule, C., Seguin, A., Retzel, E., Butterfield, Y., Barber, S., Yang, G., Stott, J., Sidiqi, A., Holt, R., Maïra, M. and Mackay, J.
Arborea EST sequencing in Picea glauca (white spruce)
Unpublished (2004)
JOURNAL
TITLE Contact: John Mackay
COMMENT Centre de Recherche en Biologie Forestiere
Universite Laval
Pavillon Charles-Eugene Marchand, Quebec, CANADA G1K 7P4
Fax: 418 656 7493
Email: jmackay@rsvs.ulaval.ca
Center for Computational Genomics and Bioinformatics (CCGB),
University of Minnesota, MN Id Identifier: MN5197337 Clone ID:
GQ0081.BR.D17 Clones available through: John Mackay, Ph. D.
Professeur adjoint -Assistant professeur EMAIL:
jmackay@rsvs.ulaval.ca Centre de Recherche en Biologie Forestiere
(Forest Biology Research Center) Universite Laval Quebec, Quebec
CANADA G1K 7P4
Plate: 1.BR row: 17 column: D
Seq primer: M3 Reverse Primer.
Location/Qualifiers

FEATURES

source
1..774
/organism="Picea glauca"
/mol_type="mRNA"
/strain="pg-653"
/db_xref="taxon:3330"
/clone="GQ0081.BR.D17"
/sex="Hermaphrodite"
/tissue_type="Vascular cambium, secondary phloem and bark of trees girdled by removing a ring of bark ca. 1 cm wide from the midpoint of the main stem"
/dev_stage="Vascular cambium, secondary phloem, and bark from trees harvested 1 and 7 days after girdling treatment"
/lab_host="E. coli DH10B cells"
/clone_1lb="GQ008: Cambium, phloem and bark of girdled saplings"
/note="Organ: Main stem of 4 year old saplings, approximately 60 cm tall; Vector: pBluescript II SK (+) XR; Site 1: Eco-RI; Site 2: Xho-I; Tissues were pooled from above and below the girdle. cDNA was prepared from 5 mg of poly A+ selected RNA and was directionally ligated into the pBluescript II SK (+) XR vector (Stratagene), transformed by electroporation into DH10B cells (in vitro) for propagation"

ORIGIN

Query Match 100.0%; Score 25; DB 7; Length 774;
Best Local Similarity 100.0%; Pred. No. 0.68;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGAGCCCTCTCGAGAGTTCCGCT 25
|||||
Db 687 CTGAGCCCTCTCGAGAGTTCCGCT 711

RESULT 11
LOCUS CO472728 805 bp mRNA linear EST 09-JUL-2004
DEFINITION GQ0081.BR.1_P04 GQ008: Cambium, phloem and bark of girdled saplings
ACCESSION CO472728
VERSION CO472728.1 GI:50141235
KEYWORDS EST
SOURCE Picea glauca (white spruce)
ORGANISM Picea glauca
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
1 (bases 1 to 805)
Morency,M.-J., Cooke,J., Pavy,N., Parsons,L., Paule,C., Seguin,A.,
Retzel,E., Butterfield,Y., Barber,S., Yang,G., Stott,J.,
Siddiqui,A., Holt,R., Marra,M. and Mackay,J.
Arborea EST sequencing in Picea glauca (white spruce)
Unpublished (2004)
Contact: John Mackay
Centre de Recherche en Biologie Forestiere
Universite Laval
Pavillon Charles-Eugene Marchand, Quebec, Quebec, CANADA G1K 7P4
Fax: 418 656 7493
Email: jmackay@rsvs.ulaval.ca
Center for Computational Genomics and Bioinformatics (CCGB),
University of Minnesota, MN id Identifier: MNS198011 Clone ID:
GQ0081.BR.P04 Clones available through: John Mackay, Ph. D.
Professeur adjoint - Assistant professeur EMAIL:
jmackay@rsvs.ulaval.ca Centre de Recherche en Biologie Forestiere
(Forest Biology Research Center) Universite Laval Quebec, Quebec
CANADA G1K 7P4
Plate: 1.BR row: 04 column: P
Seq primer: M13 Reverse Primer.
Location/Qualifiers
1. 805
/organism="Picea glauca"
/mol_type="mRNA"
/strain="pg-653"
/db_xref="taxon:3330"
/clone="GQ0081.BR_P04"
/sex="Hermaphrodite"
/tissue_type="Vascular cambium, secondary phloem and bark
of trees girdled by removing a ring of bark ca. 1 cm wide
from the midpoint of the main stem"
/dev_stage="Vascular cambium, secondary phloem, and bark
from trees harvested 1 and 7 days after girdling
treatment"
/lab_host="E. coli DH10B cells"
/clone_lib="GQ008: Cambium, phloem and bark of girdled
saplings"
/note="Organ: Main stem of 4 year old saplings,
approximately 60 cm tall; Vector: plasmid II SK (+)
XR; Site_1: Eco-RI; Site_2: Xho-I; Tissues were pooled
from above and below the girdle. cDNA was prepared from 5
mg of poly A+ selected RNA and was directionally ligated
into the plasmid II SK (+) XR vector (Stratagene),
transformed by electroporation into DH10B cells (in
vitrogen) for propagation"

ORIGIN

Query Match 100.0%; Score 25; DB 7; Length 805;
Best Local Similarity 100.0%; Pred. No. 0.69;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGAGCCCTCTCGAGAGTTCCGCT 25
|||||
Db 695 CTGAGCCCTCTCGAGAGTTCCGCT 719

RESULT 12
LOCUS CO472729 817 bp mRNA linear EST 09-JUL-2004
DEFINITION GQ0081.BR.1_E24 GQ008: Cambium, phloem and bark of girdled saplings
ACCESSION CO472729
VERSION CO472729.1 GI:50141237
KEYWORDS EST
SOURCE Picea glauca (white spruce)
ORGANISM Picea glauca
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
1 (bases 1 to 817)
Morency,M.-J., Cooke,J., Pavy,N., Parsons,L., Paule,C., Seguin,A.,
Retzel,E., Butterfield,Y., Barber,S., Yang,G., Stott,J.,
Siddiqui,A., Holt,R., Marra,M. and Mackay,J.
Arborea EST sequencing in Picea glauca (white spruce)
Unpublished (2004)
Contact: John Mackay
Centre de Recherche en Biologie Forestiere
Universite Laval
Pavillon Charles-Eugene Marchand, Quebec, Quebec, CANADA G1K 7P4
Fax: 418 656 7493
Email: jmackay@rsvs.ulaval.ca
Center for Computational Genomics and Bioinformatics (CCGB),
University of Minnesota, MN id Identifier: MNS198011 Clone ID:
GQ0081.BR.E24 Clones available through: John Mackay, Ph. D.
Professeur adjoint - Assistant professeur EMAIL:
jmackay@rsvs.ulaval.ca Centre de Recherche en Biologie Forestiere
(Forest Biology Research Center) Universite Laval Quebec, Quebec
CANADA G1K 7P4
Plate: 1.BR row: 24 column: E
Seq primer: M13 Reverse Primer.
Location/Qualifiers
1. 817
/organism="Picea glauca"
/mol_type="mRNA"
/strain="pg-653"
/db_xref="taxon:3330"
/clone="GQ0081.BR_E24"
/sex="Hermaphrodite"
/tissue_type="Vascular cambium, secondary phloem and bark
of trees girdled by removing a ring of bark ca. 1 cm wide
from the midpoint of the main stem"
/dev_stage="Vascular cambium, secondary phloem, and bark
from trees harvested 1 and 7 days after girdling
treatment"
/lab_host="E. coli DH10B cells"
/clone_lib="GQ008: Cambium, phloem and bark of girdled
saplings"
/note="Organ: Main stem of 4 year old saplings,
approximately 60 cm tall; Vector: plasmid II SK (+)
XR; Site_1: Eco-RI; Site_2: Xho-I; Tissues were pooled
from above and below the girdle. cDNA was prepared from 5
mg of poly A+ selected RNA and was directionally ligated
into the plasmid II SK (+) XR vector (Stratagene),
transformed by electroporation into DH10B cells (in
vitrogen) for propagation"

ORIGIN

Query Match 100.0%; Score 25; DB 7; Length 817;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGAGCCCTCTCGAGAGTTCCGCT 25
|||||

Db 687 CTGACCCCTCTCGAGAGTTCGCT 711

RESULT 13

LOCUS CO472694

DEFINITION GQ0081.BR.1 J17 GQ008: Cambium, phloem and bark of girdled saplings

ACCESSION Picea glauca cDNA clone GQ0081.BR_J17 5', mRNA sequence.

CO472694 819 bp mRNA linear EST 09-JUL-2004

VERSION CO472694.1 GI:50141166

KEYWORDS EST.

SOURCE Picea glauca (white spruce)

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.

REFERENCE 1 (bases 1 to 819)

AUTHORS Morency,M.-J., Cooke,J., Pavy,N., Parsons,L., Paule,C., Seguin,A., Retzel,E., Butterfield,Y., Barber,S., Yang,G., Stott,J., Siddiqui,A., Holt,R., Marra,M. and Mackay,J.

TITLE Arborea EST sequencing in Picea glauca (white spruce)

JOURNAL Unpublished (2004)

COMMENT Contact: John Mackay
Centre de Recherche en Biologie Forestiere
Universite Laval
Pavillon Charles-Eugene Marchand, Quebec, Quebec, CANADA G1K 7P4
Fax: 418 656 7493
Email: jmackay@rsvs.ulaval.ca
Center for Computational Genomics and Bioinformatics (CCGB),
University of Minnesota, MN id Identifier: MN5197649 Clone ID:
GQ0081.BR.J17 Clones available through: John Mackay, Ph. D.
Professeur adjoint - Assistant professor EMAIL:
jmackay@rsvs.ulaval.ca Centre de Recherche en Biologie Forestiere
(Forest Biology Research Center) Universite Laval Quebec, Quebec
CANADA G1K 7P4
Plate: 1.BR row: 17 column: J
Seq primer: M3 Reverse Primer.

FEATURES

source

1. 819

Location/Qualifiers

/organism="Picea glauca"

/mol_type="mRNA"

/strain="pg-653"

/db_xref="taxon:3330"

/clone="GQ0081.BR.J17"

/sex="Hermaphrodite"

/tissue_type="Vascular cambium, secondary phloem and bark of trees girdled by removing a ring of bark ca. 1 cm wide from the midpoint of the main stem"

/dev_stage="Vascular cambium, secondary phloem, and bark from trees harvested 1 and 7 days after girdling treatment"

/lab_host="E. coli DH10B cells"

/clone_1fb="GQ008: Cambium, phloem and bark of girdled saplings"

/note="Organ: Main stem of 4 year old saplings, approximately 60 cm tall; Vector: pBluescript II SK (+) XR; Site_1: Eco-RI; Site_2: Xho-I; Tissues were pooled from above and below the girdle. cDNA was prepared from 5 mg of poly A+ selected RNA and was directionally ligated into the pBluescript II SK (+) XR vector (Stratagene), transformed by electroporation into DH10B cells (in vitrogen) for propagation"

ORIGIN

Query Match 100.0%; Score 25; DB 7; Length 819;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGACCCCTCTCGAGAGTTCGCT 25

|||||

Db 695 CTGACCCCTCTCGAGAGTTCGCT 719

|||||

RESULT 14

CO472674

LOCUS CO472674

DEFINITION GQ0081.BR.1 B11 GQ008: Cambium, phloem and bark of girdled saplings

ACCESSION Picea glauca cDNA clone GQ0081.BR_B11 5', mRNA sequence.

CO472674 820 bp mRNA linear EST 09-JUL-2004

VERSION CO472674.1 GI:50141124

KEYWORDS EST.

SOURCE Picea glauca (white spruce)

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.

REFERENCE 1 (bases 1 to 820)

AUTHORS Morency,M.-J., Cooke,J., Pavy,N., Parsons,L., Paule,C., Seguin,A., Retzel,E., Butterfield,Y., Barber,S., Yang,G., Stott,J., Siddiqui,A., Holt,R., Marra,M. and Mackay,J.

TITLE Arborea EST sequencing in Picea glauca (white spruce)

JOURNAL Unpublished (2004)

COMMENT Contact: John Mackay
Centre de Recherche en Biologie Forestiere
Universite Laval
Pavillon Charles-Eugene Marchand, Quebec, Quebec, CANADA G1K 7P4
Fax: 418 656 7493
Email: jmackay@rsvs.ulaval.ca
Center for Computational Genomics and Bioinformatics (CCGB),
University of Minnesota, MN id Identifier: MN5197207 Clone ID:
GQ0081.BR.B11 Clones available through: John Mackay, Ph. D.
Professeur adjoint - Assistant professor EMAIL:
jmackay@rsvs.ulaval.ca Centre de Recherche en Biologie Forestiere
(Forest Biology Research Center) Universite Laval Quebec, Quebec
CANADA G1K 7P4
Plate: 1.BR row: 11 column: B
Seq primer: M3 Reverse Primer.

FEATURES

source

1. 820

Location/Qualifiers

/organism="Picea glauca"

/mol_type="mRNA"

/strain="pg-653"

/db_xref="taxon:3330"

/clone="GQ0081.BR.B11"

/sex="Hermaphrodite"

/tissue_type="Vascular cambium, secondary phloem and bark of trees girdled by removing a ring of bark ca. 1 cm wide from the midpoint of the main stem"

/dev_stage="Vascular cambium, secondary phloem, and bark from trees harvested 1 and 7 days after girdling treatment"

/lab_host="E. coli DH10B cells"

/clone_1fb="GQ008: Cambium, phloem and bark of girdled saplings"

/note="Organ: Main stem of 4 year old saplings, approximately 60 cm tall; Vector: pBluescript II SK (+) XR; Site_1: Eco-RI; Site_2: Xho-I; Tissues were pooled from above and below the girdle. cDNA was prepared from 5 mg of poly A+ selected RNA and was directionally ligated into the pBluescript II SK (+) XR vector (Stratagene), transformed by electroporation into DH10B cells (in vitrogen) for propagation"

ORIGIN

Query Match 100.0%; Score 25; DB 7; Length 820;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGACCCCTCTCGAGAGTTCGCT 25

|||||

Db 695 CTGACCCCTCTCGAGAGTTCGCT 719

|||||

RESULT 15

LOCUS CO472692

DEFINITION GQ0081.BR.1 J08 GQ008: Cambium, phloem and bark of girdled saplings

ACCESSION Picea glauca cDNA clone GQ0081.BR_J08 5', mRNA sequence.

CO472692 825 bp mRNA linear EST 09-JUL-2004

VERSION CO472692

KEYWORDS EST.

SOURCE Picea glauca (white spruce)

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.

REFERENCE 1 (bases 1 to 825)

AUTHORS Morency,M.-J., Cooke,J., Pavy,N., Parsons,L., Paule,C., Seguin,A., Retzel,E., Butterfield,Y., Barber,S., Yang,G., Stott,J., Siddiqui,A., Holt,R., Marra,M. and Mackay,J.

TITLE Arborea EST sequencing in Picea glauca (white spruce)

JOURNAL Unpublished (2004)

COMMENT Contact: John Mackay
Centre de Recherche en Biologie Forestiere
Universite Laval
Pavillon Charles-Eugene Marchand, Quebec, Quebec, CANADA G1K 7P4
Fax: 418 656 7493
Email: jmackay@rsvs.ulaval.ca
Center for Computational Genomics and Bioinformatics (CCGB),
University of Minnesota, MN id Identifier: MN5197207 Clone ID:
GQ0081.BR.J08 Clones available through: John Mackay, Ph. D.
Professeur adjoint - Assistant professor EMAIL:
jmackay@rsvs.ulaval.ca Centre de Recherche en Biologie Forestiere
(Forest Biology Research Center) Universite Laval Quebec, Quebec
CANADA G1K 7P4
Plate: 1.BR row: 11 column: B
Seq primer: M3 Reverse Primer.

FEATURES

source

1. 825

Location/Qualifiers

/organism="Picea glauca"

/mol_type="mRNA"

/strain="pg-653"

/db_xref="taxon:3330"

/clone="GQ0081.BR.B11"

/sex="Hermaphrodite"

/tissue_type="Vascular cambium, secondary phloem and bark of trees girdled by removing a ring of bark ca. 1 cm wide from the midpoint of the main stem"

/dev_stage="Vascular cambium, secondary phloem, and bark from trees harvested 1 and 7 days after girdling treatment"

/lab_host="E. coli DH10B cells"

/clone_1fb="GQ008: Cambium, phloem and bark of girdled saplings"

/note="Organ: Main stem of 4 year old saplings, approximately 60 cm tall; Vector: pBluescript II SK (+) XR; Site_1: Eco-RI; Site_2: Xho-I; Tissues were pooled from above and below the girdle. cDNA was prepared from 5 mg of poly A+ selected RNA and was directionally ligated into the pBluescript II SK (+) XR vector (Stratagene), transformed by electroporation into DH10B cells (in vitrogen) for propagation"

ORIGIN

Query Match 100.0%; Score 25; DB 7; Length 820;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGACCCCTCTCGAGAGTTCGCT 25

|||||

Db 695 CTGACCCCTCTCGAGAGTTCGCT 719

|||||

RESULT 15

ACCESSION CO472692
 VERSION CO472692.1 GI:50141162
 KEYWORDS
 SOURCE Picea glauca (white spruce)
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Centre de Recherche en Biologie Forestiere
 Universite Laval
 Pavillon Charles-Eugene Marchand, Quebec, Quebec, CANADA G1K 7P4
 Fax: 418 656 7493
 Email: jmackay@rsvs.ulaval.ca
 Center for Computational Genomics and Bioinformatics (CCGB),
 University of Minnesota, MN id Identifier: MN5197621 Clone ID:
 GQ0081.BR_J08 Clones available through: John Mackay, Ph. D.
 Professeur adjoint -Assistant professor EMAIL:
 jmackay@rsvs.ulaval.ca Centre de Recherche en Biologie Forestiere
 (Forest Biology Research Center) Universite Laval Quebec, Quebec
 CANADA G1K 7P4
 Plate: 1.BR row: 08 column: J
 Seq primer: M13 Reverse primer.
 Location/Qualifiers

FEATURES

Source

1. 825
 /organism="Picea glauca"
 /mol_type="rRNA"
 /strain="pg-553"
 /db_xref="taxon:3330"
 /clone="GQ0081.BR_J08"
 /sex="Hermaphrodite"
 /tissue_type="Vascular cambium, secondary phloem and bark
 of trees girdled by removing a ring of bark ca. 1 cm wide
 from the midpoint of the main stem"
 /dev_stage="Vascular cambium, secondary phloem, and bark
 from trees harvested 1 and 7 days after girdling
 treatment"
 /lab_host="E. coli DH10B cells"
 /clone_id="GQ008: Cambium, phloem and bark of girdled
 saplings"
 /note="Organ: Main stem of 4 year old saplings,
 approximately 60 cm tall; Vector: pBluescript II SK (+)
 XR; Site_1: Eco-RI; Site_2: Xho-I; Tissues were pooled
 from above and below the girdle. cDNA was prepared from 5
 mg of poly A+ selected RNA and was directionally ligated
 into the pBluescript II SK (+) XR vector (Stratagene),
 transformed by electroporation into DH10B cells (in
 vitro) for propagation"

ORIGIN

Query Match 100.0%; Score 25; DB 7; Length 825;
 Best Local Similarity 100.0%; Pred. No. 0.69;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTGAGCCCTCTCTCGAGATTCCGCT 25
 |||||
 DB 695 CTGAGCCCTCTCTCGAGATTCCGCT 719
 |||||

Search completed: March 1, 2005, 01:58:18
 Job time: 1738.75 secs

CC e.g. to increase resistance against insects or herbicides
XX
SQ Sequence 22 BP; 8 A; 5 C; 5 G; 4 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 4; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACACGTAGAAATGATCATCAC 22
1 GACACGTAGAAATGATCATCAC 22

RESULT 2
AABD24070
ID AABD24070 standard; DNA; 22 BP.

XX AAD24070;
XX
XX
XX 09-APR-2002 (first entry)
XX
XX
XX Pea lectin gene Psi element.

XX Gene expression; maize; ubiquitin promoter; Ubi-1; HSE;
XX heat shock element; agronomic gene; pea; lectin gene; Psi element; ds.
XX
XX
XX Pisum sativum.

OS
XX
XX WO200194394-A2.
XX
XX
XX 13-DEC-2001.

XX 08-JUN-2001; 2001WO-US018689.

XX 09-JUN-2000; 2000US-00590558.

XX (PROD-) PRODIGENE INC.

XX Jilka JM, Hood BE, Howard JA,

XX WPI; 2002-122117/16.

XX New promoter sequences for causing expression of a structural gene
XX especially agronomic gene or open reading frame in a plant cell,
XX comprises engineered versions of the maize ubiquitin promoter.

PS Claim 11; Page 55; 68pp; English.

XX The invention relates to a promoter sequence capable of directing
XX expression of a nucleotide sequence in a plant cell comprising maize
XX ubiquitin (Ubi-1) promoter sequence with a modification so that it does
XX not include two overlapping heat shock elements (HSE) or it directs
XX expression to increase the endosperm/embryo expression ratio of the
XX protein when compared to the ratio from a wild-type ubiquitin promoter.
XX The modified Ubi-1 promoter comprises a deletion of 3', 5' or both HSEs.
XX two non-overlapping/adjacent HSEs, replacement of HSEs with a trimer of a
XX seed specific element from the promoter of pea lectin gene Psi, or
XX insertion of a transcription factor binding site in the HSE region. An
XX expression construct comprising modified Ubi-1 promoter is useful for
XX causing expression of a structural gene (agronomic genes) or open reading
XX frame in a plant cell. The modified Ubi-1 promoter increases expression
XX levels beyond those observed with native ubiquitin promoter. The present
XX sequence is pea lectin gene Psi element used in the promoter of the
XX invention

XX Sequence 22 BP; 8 A; 5 C; 5 G; 4 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 6; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.24;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACACGTAGAAATGATCATCAC 22
1 GACACGTAGAAATGATCATCAC 22

DB 1 GACACGTAGAAATGATCATCAC 22

RESULT 3
AABD32968_1
Continuation (2 of 8) of ABD32968 from base 100001 (Human cancer-associated genomic DNA)
WP Sequence split into 8 fragments LOCUS ABD32968 Accession Abd32968

WP	Fragment Name	Begin	End
WP	ABD32968_0	1	110000
WP	ABD32968_1	100001	210000
WP	ABD32968_2	200001	310000
WP	ABD32968_3	300001	410000
WP	ABD32968_4	400001	510000
WP	ABD32968_5	500001	610000
WP	ABD32968_6	600001	710000
WP	ABD32968_7	700001	779603

Query Match 85.5%; Score 18.8; DB 13; Length 110000;
Best Local Similarity 90.9%; Pred. No. 29;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GACACGTAGAAATGATCATCAC 22
1 GACACGTAGAAATGATCATCAC 22

DB 42953 GACACGTAGAAATGATCATCAC 42954

RESULT 4
AA120672

ID AA120672 standard; DNA; 142 BP.

XX AA120672;

XX 12-OCT-2001 (first entry)

XX Probe #10605 for gene expression analysis in human cervical cell sample.

XX Probe; human; microarray; gene expression; cervical epithelial cell;

XX cervical cancer; ss.

XX Homo sapiens.

XX WO200157278-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000670.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing

XX gene expression in human cervical epithelial cells.

XX Claim 25; SEQ ID NO 10605; 487bp; English.

XX The present invention relates to human single exon nucleic acid probes
XX (SNP). The present sequence is one such probe. The SNPs are derived
XX from human Hela cells. The SNPs can be used to produce a single exon
XX microarray, which can be used for measuring human gene expression in a
XX sample derived from human cervical epithelial cells. By measuring gene
XX expression, the probes are therefore useful in grading and/or staging of
XX diseases of the cervix, notably cervical cancer. Note: The sequence data

CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 142 BP; 52 A; 40 C; 20 G; 30 T; 0 U; 0 Other;

Query Match 78.2%; Score 17.2; DB 4; Length 142;

Best Local Similarity 86.4%; Pred. No. 86;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GACACGTAGATGAGTCATCAC 22

DB 80 GACACATAAATCAGTCATCAC 101

RESULT 5

ABA65716
ID ABA65716 standard; DNA; 142 BP.

XX ABA65716;

DT 01-FEB-2002 (first entry)

XX Human foetal liver single exon nucleic acid probe #14021.

XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

OS Homo sapiens.

PN WO200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000669.

PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-483447/52.

PT Human genome-derived single exon nucleic acid probes useful for analyzing

XX gene expression in human fetal liver.

PS Claim 4; SEQ ID NO 14021; 639pp + Sequence Listing; English.

XX The invention relates to a single exon nucleic acid probe for measuring

CC human gene expression in a sample derived from human foetal liver. The

CC single exon nucleic acid probes may be used for predicting, measuring and

CC displaying gene expression in samples derived from human fetal liver. The

CC present sequence is a single exon nucleic acid probe of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 142 BP; 52 A; 40 C; 20 G; 30 T; 0 U; 0 Other;

XX Query Match 78.2%; Score 17.2; DB 4; Length 142;

XX Best Local Similarity 86.4%; Pred. No. 86;

XX Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GACACGTAGATGAGTCATCAC 22

DB 80 GACACATAAATCAGTCATCAC 101

RESULT 6
AA145884
ID AA145884 standard; DNA; 142 BP.

XX AA145884;

DT 17-OCT-2001 (first entry)

XX Probe #14570 used to measure gene expression in human placenta sample.

XX Probe; microarray; human; placenta; antenatal diagnosis;

XX genetic disorder; ss.

OS Homo sapiens.

PN WO200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000663.

PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488897/53.

PT Human genome-derived single exon nucleic acid probes useful for analyzing

XX gene expression in human placenta.

PS Claim 25; SEQ ID NO 14570; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP).

CC The present sequence is one such probe. The probes are useful for

CC producing a microarray for predicting, measuring and displaying gene

CC expression in samples derived from human placenta. The probes are useful

CC for antenatal diagnosis of human genetic disorders

XX Sequence 142 BP; 52 A; 40 C; 20 G; 30 T; 0 U; 0 Other;

XX Query Match 78.2%; Score 17.2; DB 4; Length 142;

XX Best Local Similarity 86.4%; Pred. No. 86;

XX Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GACACGTAGATGAGTCATCAC 22

DB 80 GACACATAAATCAGTCATCAC 101

RESULT 7
ABA32803
ID ABA32803 standard; DNA; 142 BP.

XX ABA32803;

DT 23-JAN-2002 (first entry)

XX Probe #11269 for gene expression analysis in human heart cell sample.

XX Human; gene expression; heart; microarray; vascular system; probe;

XX cardiovascular disease; hypertension; cardiac arrhythmia;

XX congenital heart disease; ss.

OS Homo sapiens.

```
XX PN WO200157274-A2.
XX PD 09-AUG-2001.
XX PF 30-JUN-2001; 2001WO-US000666.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-0207456P.
XX PR 03-AUG-2000; 2000US-00608408.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488990/53.
XX DR WPI; 2001-488990/53.
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX PT hearts.
XX PS Claim 4; SEQ ID NO 11269; 530bp; English.
XX SQ The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease. Note: The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_sequences
XX SQ Sequence 142 BP; 52 A; 40 C; 20 G; 30 T; 0 U; 0 Other;
Query Match 78.2%; Score 17.2; DB 4; Length 142;
Best Local Similarity 86.4%; Pred. No. 86;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GACACGTAGATGATCATCAC 22
Db 80 GACACATTAATCATCATCATC 101
RESULT 8
AAK39862
ID AAK39862 standard; DNA; 142 BP.
XX AAK39862;
XX AC
XX DT 06-NOV-2001 (first entry)
XX DT
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 14419.
XX DE
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukemia; lymphoma; myeloma; ss.
XX OS Homo sapiens.
XX OS
XX PN WO200157276-A2.
XX PN
XX PD 09-AUG-2001.
XX PD
XX PF 30-JAN-2001; 2001WO-US000668.
XX PF
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
```

```
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488990/53.
XX DR WPI; 2001-488990/53.
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human bone marrow.
XX PS Example 4; SEQ ID NO 14419; 658bp + Sequence Listing; English.
XX SQ The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukemia and myeloma. The present sequence is one of
CC the probes of the invention
XX SQ Sequence 142 BP; 52 A; 40 C; 20 G; 30 T; 0 U; 0 Other;
Query Match 78.2%; Score 17.2; DB 4; Length 142;
Best Local Similarity 86.4%; Pred. No. 86;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GACACGTAGATGATCATCAC 22
Db 80 GACACATTAATCATCATCATC 101
RESULT 9
AAK14121
ID AAK14121 standard; DNA; 142 BP.
XX AAK14121;
XX AC
XX DT 05-NOV-2001 (first entry)
XX DT
XX DE Human brain expressed single exon probe SEQ ID NO: 14112.
XX DE
XX KW Human; brain expressed exon; gene expression analysis; probe; microarray;
XX KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;
XX KW ss.
XX OS Homo sapiens.
XX OS
XX PN WO200157275-A2.
XX PN
XX PD 09-AUG-2001.
XX PD
XX PF 30-JAN-2001; 2001WO-US000667.
XX PF
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-483446/52.
XX DR WPI; 2001-483446/52.
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX PT brains.
```

PS Example 4; SEQ ID NO 14112; 650bp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention
XX
SQ Sequence 142 BP; 52 A; 40 C; 20 G; 30 T; 0 U; 0 Other;
XX
Query Match 78.2%; Score 17.2; DB 4; Length 142;
Best Local Similarity 86.4%; Pred. No. 86;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
OY 1 GACACGTAGATGAGTCATCAC 22
DB 80 GACACATAAATCAGTCATCAC 101
XX
RESULT 10
ABS39453
ID ABS39453 standard; DNA; 142 BP.
XX
AC ABS39453;
XX
DT 25-FEB-2003 (first entry)
XX
DE Human liver single exon probe, SEQ ID NO 14443.
XX
KW Human; single exon nucleic acid probe; liver; cirrhosis;
KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
KW coronary heart disease; ss.
XX
OS Homo sapiens.
XX
PN MO200157273-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000664.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488898/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human adult liver.
XX
PS Claim 4; SEQ ID NO 14443; 658bp; English.
XX
CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/fragments). The probe hybridizes at high
CC stringency to a nucleic acid molecule expressed in the human adult liver.
CC (I) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. ABS25011-AB851005 represent human
CC liver single exon nucleic acid probes of the invention. Note: The

CC sequence information for this patent does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 142 BP; 52 A; 40 C; 20 G; 30 T; 0 U; 0 Other;
XX
Query Match 78.2%; Score 17.2; DB 4; Length 142;
Best Local Similarity 86.4%; Pred. No. 86;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
OY 1 GACACGTAGATGAGTCATCAC 22
DB 80 GACACATAAATCAGTCATCAC 101
XX
RESULT 11
ABS13960
ID ABS13960 standard; DNA; 142 BP.
XX
AC ABS13960;
XX
DT 19-AUG-2002 (first entry)
XX
DE Human genome-derived single exon probe ORF from lung SEQ ID NO 13951.
XX
KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemangioma;
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karsenger syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease; open reading frame; ORF.
XX
OS Homo sapiens.
XX
PN MO200186003-A2.
XX
PD 15-NOV-2001.
XX
PF 30-JAN-2001; 2001WO-US000665.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2002-114183/15.
XX
PT Spatially-addressable set of single exon nucleic acid probes, used to
XX measure gene expression in human lung samples.
XX
PS Claim 4; SEQ ID NO 13951; 634bp; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of probes
CC; the novel set of probes which hybridize at high stringency to a nucleic
CC acid expressed in the human lung; measuring gene expression in a sample
CC derived from human lung; comprising (a) contacting the array with a
CC collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of the

CC array; identifying exons in a eukaryotic genome, comprising (a)
 CC algorithmically predicting at least one exon from genomic sequences of
 CC the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene expression
 CC analysis, and for identifying exons in a gene, particularly using human
 CC lung derived mRNA and for the study of lung diseases such as asthma, lung
 CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
 CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
 CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
 CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
 CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
 CC Karsenger syndrome, fibrocystic pulmonary dysplasia, primary ciliary
 CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The
 CC present sequence is a single exon probe open reading frame of the
 CC invention. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX

Query Match 78.2%; Score 17.2; DB 6; Length 142;
 Best Local Similarity 86.4%; Pred. No. 86;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GACACGTAGATGATCATCAC 22
 |||||
 Db 80 GACACATTAATCATCATCAC 101

RESULT 12

AA11458
 ID AA11458 standard; DNA; 483 BP.

AA11458;

12-OCT-2001 (first entry)

Probe #1391 for gene expression analysis in human cervical cell sample.

Probe; human; microarray; gene expression; cervical epithelial cell;
 cervical cancer; ss.

Homo sapiens.

WO200157278-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US000670.

04-FEB-2000; 2000US-0180312P.

26-MAY-2000; 2000US-0207456P.

30-JUN-2000; 2000US-00608408.

03-AUG-2000; 2000US-00632366.

21-SEP-2000; 2000US-0234687P.

27-SEP-2000; 2000US-0236359P.

04-OCT-2000; 2000GB-00024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI, 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing
 PT gene expression in human cervical epithelial cells.
 PT
 XX
 PS Claim 25; SEQ ID NO 1391; 487bp; English.

CC The present invention relates to human single exon nucleic acid probes
 CC (SENPs). The present sequence is one such probe. The SENPs are derived
 CC from human HeLa cells. The SENPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a
 CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging of
 CC diseases of the cervix, notably cervical cancer. Note: The sequence data
 CC for this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX

Sequence 483 BP; 151 A; 106 C; 99 G; 127 T; 0 U; 0 Other;

Query Match 78.2%; Score 17.2; DB 4; Length 483;
 Best Local Similarity 86.4%; Pred. No. 1e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GACACGTAGATGATCATCAC 22
 |||||
 Db 402 GACACATTAATCATCATCAC 423

RESULT 13

ABA53123
 ID ABA53123 standard; DNA; 483 BP.

ABA53123;

01-FEB-2002 (first entry)

Human foetal liver single exon nucleic acid probe #1428.

Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

Homo sapiens.

WO200157277-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US000669.

04-FEB-2000; 2000US-0180312P.

26-MAY-2000; 2000US-0207456P.

30-JUN-2000; 2000US-00608408.

03-AUG-2000; 2000US-00632366.

21-SEP-2000; 2000US-0234687P.

27-SEP-2000; 2000US-0236359P.

04-OCT-2000; 2000GB-00024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI, 2001-483447/52.

Human genome-derived single exon nucleic acid probes useful for analyzing
 PT gene expression in human foetal liver.
 PT
 XX
 PS Claim 1; SEQ ID NO 1428; 639bp + Sequence Listing; English.

CC The invention relates to a single exon nucleic acid probe for measuring
 CC human gene expression in a sample derived from human foetal liver. The
 CC single exon nucleic acid probes may be used for predicting, measuring and
 CC displaying gene expression in samples derived from human foetal liver. The
 CC present sequence is a single exon nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 483 BP; 151 A; 106 C; 99 G; 127 T; 0 U; 0 Other;

Query Match 78.2%; Score 17.2; DB 4; Length 483;
Best Local Similarity 86.4%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GACACGTAGATGACTCATCAC 22
DB 402 GACACATAAATCAGTCATCAC 423

RESULT 14

AA132730
ID AA132730 standard; DNA; 483 BP.

AA132730;

DT 17-OCT-2001 (first entry)

DE Probe #1416 used to measure gene expression in human placenta sample.

XX Probe; microarray; human; placenta; antenatal diagnosis;
KM genetic disorder; ss.

OS Homo sapiens.

PN WO200157274-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000663.

PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488897/53.

PT Human genome-derived single exon nucleic acid probes useful for analyzing

PT gene expression in human placenta.

PS Claim 25; SEQ ID NO 1416; 654bp; English.

CC The present invention relates to single exon nucleic acid probes (SENP).

CC The present sequence is one such probe. The probes are useful for

CC producing a microarray for predicting, measuring and displaying gene

CC expression in samples derived from human placenta. The probes are useful

CC for antenatal diagnosis of human genetic disorders

RESULT 15

ABA22897

ID ABA22897 standard; DNA; 483 BP.

XX ABA22897;

DT 23-JAN-2002 (first entry)

DE Probe #1363 for gene expression analysis in human heart cell sample.

XX Human; gene expression; heart; microarray; vascular system; probe;

KM cardiovascular disease; hypertension; cardiac arrhythmia;

KW congenital heart disease; ss.

OS Homo sapiens.

PN WO200157274-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000666.

PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488897/53.

PT Single exon nucleic acid probes for analyzing gene expression in human

PT hearts.

PS Claim 1, SEQ ID NO 1363; 530bp; English.

CC The present invention relates to single exon nucleic acid probes for

CC measuring human gene expression in a sample derived from human heart. The

CC present sequence is one such probe. The probes may be used for

CC predicting, measuring and displaying gene expression in samples derived

CC from the human heart via microarrays. By measuring gene expression, the

CC probes are useful for predicting, diagnosing, grading, staging,

CC monitoring and prognosing diseases of the human heart and vascular system

CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and

CC congenital heart disease. Note: The sequence data for this patent did not

CC form part of the printed specification, but was obtained in electronic

CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

QY 1 GACACGTAGATGACTCATCAC 22

DB 402 GACACATAAATCAGTCATCAC 423

Search completed: February 28, 2005, 23:51:50
Job time : 198 secs

This Page Blank (uspto)

ORGANISM Agrobacterium tumefaciens
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.

REFERENCE 1
AUTHORS Jilka, J.M., Hood, E.E. and Howard, J.A.
TITLE Novel plant promoter sequences and methods of use for same
JOURNAL Patent: WO 0194394-A 5 13-DEC-2001;
Prodigene, Inc. (US)
FEATURES
source 1. .22
/organism="Agrobacterium tumefaciens"
/mol_type="unassigned DNA"
/db_xref="taxon:358"

ORIGIN
Query Match 100.0%; Score 22; DB 6; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACACGTAGATGATCATCAC 22
|||||
1 GACACGTAGATGATCATCAC 22

Db 1 GACACGTAGATGATCATCAC 22

RESULT 3
LOCUS AY192160 827 bp DNA linear SYN 26-FEB-2003
DEFINITION Synthetic construct multi-copy enhancer promoter element, complete
ACCESSION AY192160
VERSION AY192160
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
1. .1659
/organism="Pisum sativum"
/mol_type="mRNA"
/db_xref="taxon:3888"
/product="lectin mRNA"
/note="lectin"
/codon_start=1
/protein_id="AAA33676.1"
/db_xref="GI:169113"
/translation="MASLQTMISFYAFLSLTLTLFFPVNSTETTSFLITKSPD
QGNLIFQDGYTKEKTLTKAVKTVRAVLSPIHMDREKGVANFTSPFTEVIN
APNSYNVADGFTFFIAPVDTRKPTGCGYLGVSNAEDKTOJVAVEDTFYNAADP
SNRDHIGIDVNSIKSVNTKSMKLGNGEAVVIAFNATNVLTVSLTPNSLSEENY
TSYLSDDVYSIKDVPPEWVRIGFATTTAEVAAHVLVSWSPHSELSTGSSKQADA"

ORIGIN 936 bp upstream of BamHI site.

Query Match 100.0%; Score 22; DB 8; Length 1659;
Best Local Similarity 100.0%; Pred. No. 0.89;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACACGTAGATGATCATCAC 22
|||||
392 GACACGTAGATGATCATCAC 413

Db 392 GACACGTAGATGATCATCAC 413

RESULT 5
LOCUS A14134 1701 bp DNA linear PAT 15-JAN-2003
DEFINITION promoter and lectin gene.
ACCESSION A14134
VERSION A14134.1 GI:490034
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
1. .1701
/organism="Pisum sativum"
/mol_type="unassigned DNA"
/db_xref="taxon:3888"

ORIGIN

Query Match 100.0%; Score 22; DB 12; Length 827;
Best Local Similarity 100.0%; Pred. No. 0.94;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACACGTAGATGATCATCAC 22
|||||
|||||

Db 1 GACACGTAGATGATCATCAC 22

FEATURES
source
1. .827
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/note="multi-copy enhancer promoter element"
/note="contains 3 copies of 30bp element from glycine
promoter"
91. .160
/note="contains 3 copies of 22bp W1 seed-specific
cis-element from pea lectin promoter"
161. .376
/note="contains 3 copies of 68bp UNS from phaseolin
promoter"
501. .827
/note="contains CaMV mini promoter (-60) and TMV omega
enhancer"

Db 91 GACACGTAGATGATCATCAC 112

RESULT 4
PEALCEB 1659 bp mRNA linear PLN 27-APR-1993
LOCUS
DEFINITION Pea PEB1 gene encoding lectin, complete cds.
ACCESSION M18160
VERSION M18160.1 GI:169112
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
1. .1659
/organism="Pisum sativum"
/mol_type="mRNA"
/db_xref="taxon:3888"
/product="lectin mRNA"
/note="lectin"
/codon_start=1
/protein_id="AAA33676.1"
/db_xref="GI:169113"
/translation="MASLQTMISFYAFLSLTLTLFFPVNSTETTSFLITKSPD
QGNLIFQDGYTKEKTLTKAVKTVRAVLSPIHMDREKGVANFTSPFTEVIN
APNSYNVADGFTFFIAPVDTRKPTGCGYLGVSNAEDKTOJVAVEDTFYNAADP
SNRDHIGIDVNSIKSVNTKSMKLGNGEAVVIAFNATNVLTVSLTPNSLSEENY
TSYLSDDVYSIKDVPPEWVRIGFATTTAEVAAHVLVSWSPHSELSTGSSKQADA"

ORIGIN 936 bp upstream of BamHI site.

Query Match 100.0%; Score 22; DB 8; Length 1659;
Best Local Similarity 100.0%; Pred. No. 0.89;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACACGTAGATGATCATCAC 22
|||||
392 GACACGTAGATGATCATCAC 413

Db 392 GACACGTAGATGATCATCAC 413

RESULT 5
LOCUS A14134 1701 bp DNA linear PAT 15-JAN-2003
DEFINITION promoter and lectin gene.
ACCESSION A14134
VERSION A14134.1 GI:490034
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
1. .1701
/organism="Pisum sativum"
/mol_type="unassigned DNA"
/db_xref="taxon:3888"

ORIGIN

Query Match 100.0%; Score 22; DB 12; Length 827;
Best Local Similarity 100.0%; Pred. No. 0.94;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACACGTAGATGATCATCAC 22
|||||
|||||

Db 1 GACACGTAGATGATCATCAC 22

FEATURES
source
1. .827
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/note="multi-copy enhancer promoter element"
/note="contains 3 copies of 30bp element from glycine
promoter"
91. .160
/note="contains 3 copies of 22bp W1 seed-specific
cis-element from pea lectin promoter"
161. .376
/note="contains 3 copies of 68bp UNS from phaseolin
promoter"
501. .827
/note="contains CaMV mini promoter (-60) and TMV omega
enhancer"

Query Match	100.0%	Score 22	DB 6	Length 1701
Beat Local Similarity	100.0%	Pred. No. 0.89		
Matches 22	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Db	414	GACACGTGAGTGCATCATC	435	
OrigIn		<p>473. .1300 /gene="leca" 473. .1300 /gene="leca" /codon_start=1 /product="lectin" /protein_id="CAA01149.1" /db_xref="GI:490035" /translation="MASHQOMISFYALFSLILTTILFFRVNSTETTSFILTKSPDP QONILFQDGGYTKKTLITRAVKNTVGRALVSSPIHMDRETGNVAFVTSFTVIN APNSYVADGTFPIAPVDITKPGQGGYLVFNSAEYDKTQTVAVEEDFYNAADPE SNRRHIGIDVNSIKSVTKSMKLVNGEAAVVIAFNATVNLIVSLTYPNLSLEENNV TSYTLSDVNSLKDVPENWVRIGFSATTCABEYAAAEHVLSWSPHSELSTSSKQADA"</p>		
RESULT 6				
PSIECA				
LOCUS	PSIECA	1701 bp	DNA	linear
DEFINITION	Pisum sativum leca gene for lectin.			
ACCESSION	Y00440			
VERSION	Y00440.1	GI:20769		
KEYWORDS	leca gene; lectin.			
SOURCE	Pisum sativum (pea)			
ORGANISM	Pisum sativum Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; rosids; euroside I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.			
REFERENCE	1 (bases 1 to 1701) Gatehouse,J.A., Bowm,D., Evans,I.M., Gatehouse,L.N., Jones,D., Preston,P. and Croy,R.R. Sequence of the seed lectin gene from pea (Pisum sativum L.) Nucleic Acids Res. 15 (18), 7642 (1987)			
TITLE	JOURNAL			
PUBMED	88015625			
MEDLINE	3658708			
REFERENCE	2 (bases 1 to 1701) Gatehouse,J.A. Direct Submission Submitted (07-SEP-1987) Gatehouse J.A., Botany Department, University of Durham, South Road, Durham, DH1 3LE, England			
JOURNAL	University of Durham			
FEATURES	Location/Qualifiers			
source	1. .1701 /organism="Pisum sativum" /mol_type="genomic DNA" /strain="Feltbam F18c" /db_xref="taxon:3888" /tissue_type="seed" 408. .413 /note="TATA-box"			
promoter	473. .1300 /codon_start=1 /product="lectin-precursor (AA -30 to 245)" /protein_id="CAA68497.1" /db_xref="GI:20770" /db_xref="GOA:P02867" /db_xref="UniProt/Swiss-Prot:P02867" /translation="MASHQOMISFYALFSLILTTILFFRVNSTETTSFILTKSPDP QONILFQDGGYTKKTLITRAVKNTVGRALVSSPIHMDRETGNVAFVTSFTVIN APNSYVADGTFPIAPVDITKPGQGGYLVFNSAEYDKTQTVAVEEDFYNAADPE SNRRHIGIDVNSIKSVTKSMKLVNGEAAVVIAFNATVNLIVSLTYPNLSLEENNV TSYTLSDVNSLKDVPENWVRIGFSATTCABEYAAAEHVLSWSPHSELSTSSKQADA"			
CDS	473. .562 /note="signal peptide (AA -30 to -1)" 563. .1297			
sig.peptide	mat.peptide			

```

misc_feature          /product="lectin AA 1 - 245"
                      1359..1365
polyA_site            /note="polyA signal"
                      1422
polyA_site            /note="pot. polyA site"
                      1446
                      /note="alt. pot. polyA site"

ORIGIN
Query Match          100.0%; Score 22; DB 8; Length 1701;
Best Local Similarity 100.0%; Pred. No. 0.89;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Cy      1   GACACGTGATGAGTCATCC 22
        |||
Db      414 GACACGTGAAATGACTCATCAC 435

RESULT 7
PSLSELECTIN LOCUS       3360 bp     DNA         linear    PLN 31-OCT-1999
DEFINITION P.sativum psl gene for Psl lectin.
ACCESSION X66368
VERSION X66368.1 GI:562782
KEYWORDS lectin; psl gene; Psl lectin.
SOURCE Pisum sativum (pea)
ORGANISM Pisum sativum
           Eubaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots;
           rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae;
           Pisum.
REFERENCE 1 (bases 1 to 3360)
AUTHORS de Pater,S., Pham,K., Chua,N.H., Memelink,J. and Kijne,J.
TITLE A 22-bp fragment of the pea lectin promoter containing essential
JOURNAL TGAC-like motifs confers seed-specific gene expression
MEDLINE Plant Cell 5 (8), 877-886 (1993)
PUBMED 94004020
REFERENCE 2 (bases 1 to 1713)
AUTHORS Hoedemaeker,F.J., Richardson,M., Diaz,C.L., de Pater,B.S. and
JOURNAL Kijne,J.W.
TITLE Pea [Pisum sativum L.] seed isolating 1 and 2 and pea root lectin
JOURNAL result from carboxypeptidase-like processing of a single gene
MEDLINE product
PUBMED Plant Mol. Biol. 24 (1), 75-81 (1994)
JOURNAL 94154245
REFERENCE 3 (bases 1 to 3360)
AUTHORS de Pater,S., Katagiri,F., Kijne,J. and Chua,N.H.
TITLE bzfp proteins bind to a palindromic sequence without an ACCT core
JOURNAL located in a seed-specific element of the pea lectin promoter
MEDLINE Plant J. 6 (2), 133-140 (1994)
PUBMED 95004657
REFERENCE 4
JOURNAL 7920707
PUBMED
REFERENCE de Pater,B.S.
AUTHORS Direct Submission
JOURNAL Submitted (13-MAY-1992) B.S. de Pater, Center for Phytoecchnology,
TITLE Botanical Laboratory, Nomensteeg 3, 2311 VJ Leiden, THE
JOURNAL NETHERLANDS
REMARK revised by [3] MAT
REFERENCE 5 (bases 1 to 3360)
AUTHORS de Pater,B.S.
TITLE Direct Submission
JOURNAL Submitted (27-OCT-1994) B.S. de Pater, Center for Phytoecchnology,
REMARK Clueins Laboratorium, Wassenaarseweg 64, 2333 AL Leiden, THE
COMMENT NETHERLANDS
FEATURES On Nov 1, 1994 this sequence version replaced gi:20803.
source location/Qualifiers
1..3360
/organism="Pisum sativum"
/mol_type="genomic DNA"
/variety="Peldham First"

```

```

/db_xref="taxon:3888"
/clone="pP100"
protein_bind
2072..2081
/bound_motif="ASF-1"
2088..2092
(2117..2122)..>2959
2132..2959
/gene="psl"
2132..2959
/gene="psl"
/codon_start=1
/product="psl lectin"
/protein_id="CAA47011.1"
/db_xref="GI:20804"
/db_xref="GOA:P02867"
/db_xref="UniProt/Swiss-Prot:P02867"
/translation="MASLOTQWISFYALFLSLTLTLTFKKNSTETTSPLITRSPD
OONLPGCGGTTTKEKLTITKAVKNVAGALYSPDIHMDRETGVANFVSTFVFN
ARNSYVADGTFPFLAPVDTKPTQGGYAGVNSAEYDKTQTVAVEEDTPYNAADP
SRNRHIGIDVNSIKSVNKSWKLNQGEANVIAFNATVLTLYSLTPVSLSEBNY
TSYTLSDVSLNDVVEWVRIGFSAITGAEVAHEVLSMSPHSLSTSSKQADA"
2132..2221
/gene="psl"

```

sig_peptide

ORIGIN

```

Query Match      100.0%  Score 22;  DB 8;  Length 3360;
Best Local Similarity 100.0%;  Pred. No. 0.84;
Matches 22;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy      1  GACACGTAGAAATGAGTCATCAG 22
Db      2061 GACACGTAGAAATGAGTCATCAG 2082

```

```

RESULT 8
AC011143      136492 bp  DNA  linear  PRI 02-APR-2001
DEFINITION Homo sapiens chromosome 18, clone RP11-1015, complete sequence.
ACCESSION AC011143
VERSION AC011143.6 GI:13470229
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 136492)
Birten,B., Linton,L., Nusbaum,C. and Lander,E.
Unpublished
2 (bases 1 to 136492)
Birten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckertly,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Dominko,M., Donelan,L., Doyle,M.,
Farrington,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Locke,K., Kam,L., Karatas,A., Klein,J.,
Lehoczky,J., Liu,C., Locke,K., McDonald,P., Marnis,N.,
McEwan,P., McGurk,A., McKernan,K., McDonald,P., Marnis,N.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testa,D., Thirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wymen,D., Ye,W.J., Zimmer,A. and Zody,M.

```

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (01-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 136492)
Birten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Baertien,V., Boguslavsky,L., Boukhalter,B., Brown,A.,
Camarata,S., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K.,
Diaz,U.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D.,

TITLE
JOURNAL
COMMENT

Submitted (02-APR-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 28, 2001 this sequence version replaced g113324843.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu

Project Information
Center project name: L2515
Center clone name: L_O_15

FEATURES

```

source
1..136492
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-1015"
/clone_1b="RP11-11 Human Male BAC"
/complement(1250..1510)
/rpt_family="MER39"
/complement(1511..1896)
/rpt_family="THERIC"
/complement(1897..1983)
/rpt_family="MER39"
/complement(2259..2291)
/rpt_family="12"
/complement(2503..2841)
/rpt_family="THER1B"
3249..3402
/rpt_family="AluY"
3657..3954
/rpt_family="AluJ"
3956..3985
/rpt_family="AluJ"
4007..4295
/rpt_family="AluJ"
/complement(5078..5207)
/rpt_family="L1MB7"
/complement(5287..5585)
/rpt_family="L1MC4a"
5826..6242
/rpt_family="MSTB1"
/complement(6481..6807)
/rpt_family="AluSg"
7221..7394
/rpt_family="MIR"
/complement(7429..8364)
/rpt_family="LTR1B"
10152..10187
/rpt_family="(TG)n"

```

```

repeat_region complement(10614..11362)
/rpt_family="LIMC4a"
repeat_region /rpt_family="L2"
11364..11700
repeat_region /rpt_family="L2"
11903..12038
repeat_region /rpt_family="L2"
12793..12820
repeat_region /rpt_family="AT_rich"
complement(13170..13201)
repeat_region /rpt_family="U2"
complement(13287..13340)
repeat_region /rpt_family="MERSA"
13395..13443
repeat_region /rpt_family="MERSA"
complement(13401..13442)
repeat_region /rpt_family="MERSB"
14965..15007
repeat_region /rpt_family="(TCTA)n"
complement(15008..15294)
repeat_region /rpt_family="AluSx"
15296..15317
repeat_region /rpt_family="AT_rich"
15572..15759
repeat_region /rpt_family="MERSA"
complement(16050..16156)
repeat_region /rpt_family="LTR1"
complement(16219..16611)
repeat_region /rpt_family="HUBS-P2"
complement(16612..17391)
repeat_region /rpt_family="LTR1"
17546..18358
repeat_region /rpt_family="HAL1"
18486..18647
repeat_region /rpt_family="MERSB"
complement(18654..19102)
repeat_region /rpt_family="MLT1C"
19131..19273
repeat_region /rpt_family="HAL1"
19554..19621
repeat_region /rpt_family="HAL1"
19818..19845
repeat_region /rpt_family="AT_rich"
complement(20169..20627)
repeat_region /rpt_family="LIM4"
complement(20628..21367)
repeat_region /rpt_family="LIPAT"
complement(21368..21519)
repeat_region /rpt_family="LIM4"
complement(21872..22228)
repeat_region /rpt_family="MERS9"
complement(22232..22327)
repeat_region /rpt_family="LTR48B"
complement(22378..22857)
repeat_region /rpt_family="MERS9"
24790..24827
repeat_region /rpt_family="(T)n"
25131..25201
repeat_region /rpt_family="(TATATG)n"
complement(25277..26219)
repeat_region /rpt_family="LIM4"
complement(26222..26375)
repeat_region /rpt_family="LTR9"
complement(26404..26909)
repeat_region /rpt_family="LTR9"
complement(26911..34208)
repeat_region /rpt_family="HUBS-P3b"
34209..34324
repeat_region /rpt_family="(CATTC)n"
complement(34325..34399)
repeat_region /rpt_family="HUBS-P3b"
complement(34429..34947)
repeat_region /rpt_family="LTR9"
complement(34952..37378)

```

```

repeat_region /rpt_family="LIM4"
complement(37433..38509)
repeat_region /rpt_family="LIM4"
complement(38536..39436)
repeat_region /rpt_family="LIM4"
complement(39832..39969)
repeat_region /rpt_family="FLAM C"
complement(40224..40514)
repeat_region /rpt_family="AluSq"
complement(40897..41020)
repeat_region /rpt_family="MIR"
41370..41508
repeat_region /rpt_family="L2"
41513..41554
repeat_region /rpt_family="AT_rich"
complement(42256..42369)
repeat_region /rpt_family="MLT1H"

Query Match 85.5%; Score 18.8; DB 9; Length 136492;
Best Local Similarity 90.9%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GACACGTAGATGAGTCATCACC 22
Db 136431 GACACGTAGATGAGTCATCACC 136452

RESULT 9
AC090660 195156 bp DNA linear PRI 28-NOV-2001
LOCUS Homo sapiens chromosome 18, clone RP11-86901, complete sequence.
DEFINITION AC090660
AC090660.3 GI:17047149
VERSION
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Graniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 195156)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 18, clone RP11-86901
Unpublished
2 (bases 1 to 195156)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
Barra, N., Baetien, V., Boguslavsky, L., Boukhalter, B., Brown, A.,
Camata, V., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S.,
Collimore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Faro, S., Ferreira, P., Fitzhugh, M., Gage, D., Galegan, J.,
Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Karatas, A., Larocque, K., Lamazares, R., Landers, T.,
Lehoczky, V., Levine, R., Liu, G., Maclean, C., Macdonald, P.,
Marquis, N., Mathews, C., McCarthy, M., McEwan, P., McKernan, K.,
McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P.,
Sounez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Struass, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Travers, M., Travis, N., Trigilio, J., Vasilev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zainoun, J.,
Zemek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (07-MAR-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 195156)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barra, N., Baetien, V., Boguslavsky, L., Boukhalter, B.,
Brown, A., Camata, V., Campopiano, A., Chang, J., Chazaro, B.,
Choepel, Y., Colangelo, M., Collins, S., Collimore, A., Cooke, P.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,

```



```
repeat_region      31484..31525      /rpt_family="AT_rich"
repeat_region      complement(32227..32340) /rpt_family="MULTIH"
repeat_region      complement(32416..32685) /rpt_family="MULTIH"
repeat_region      complement(32765..32914) /rpt_family="MIR"
repeat_region      33795..33825      /rpt_family="(TA)n"

Query Match      85.5%; Score 18.8; DB 9; Length 195156;
Best Local Similarity 90.9%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 1 GACACGTAGATGATGATCATCAC 22
Db 126402 GACACGTAGATGATGATCACCAC 126423

RESULT 10
AP001593/c 198127 bp DNA linear HTG 26-JUL-2000
LOCUS Homo sapiens chromosome 18 clone RP11-86901 map 18q21, WORKING
DEFINITION DRAFT SEQUENCE, 11 unordered pieces.
ACCESSION AP001593
VERSION AP001593.3 GI:9501844
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 198127)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.,
Homo sapiens 198,127 genomic DNA of 18q21
Published Only in DataBase (2000)
2 (bases 1 to 198127)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.,
Direct Submission
Submitted (31-MAR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-15-1 Kitesato, Sagaminara, Kanagawa 228-8555, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:htcp://hgp.gsc.riken.go.jp/,
Tel:81-42-778-9923, Fax:81-42-778-9924)
On Jul 26, 2000 this sequence version replaced gi:8117417.

----- Genome Center
Center: RIKEN Genomic Sciences Center (GSC)
Center code: RIKEN
Web site: htcp://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDrafft18
Center clone name: RP11-86901
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 195298 bases at least Q40
Consensus quality: 196428 bases at least Q30
Consensus quality: 196782 bases at least Q20
Insert size: 197127; sum-of contigs
Quality coverage: 9.52x in Q20 bases; sum-of-contigs

-----
NOTE: This is a 'working draft' sequence. It currently consists of
11 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 58065 contig of 58065 bp in length
```

```
58166 92043 contig of 33878 bp in length
92144 111587 contig of 19444 bp in length
111688 129552 contig of 17865 bp in length
129653 149317 contig of 19665 bp in length
149418 164242 contig of 14825 bp in length
164343 175076 contig of 10734 bp in length
175177 185856 contig of 10680 bp in length
185957 195249 contig of 9223 bp in length
195350 196868 contig of 1519 bp in length
196969 198127 contig of 1159 bp in length.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 58065: contig of 58065 bp in length
* 58066 58165: gap of 100 bp
* 58166 92043: contig of 33878 bp in length
* 92044 92143: gap of 100 bp
* 92144 111587: contig of 19444 bp in length
* 111588 111687: gap of 100 bp
* 111688 129552: contig of 17865 bp in length
* 129553 129652: gap of 100 bp
* 129653 149317: contig of 19665 bp in length
* 149318 149417: gap of 100 bp
* 149418 164242: contig of 14825 bp in length
* 164243 164342: gap of 100 bp
* 164343 175076: contig of 10734 bp in length
* 175077 175176: gap of 100 bp
* 175177 185856: contig of 10680 bp in length
* 185857 185956: gap of 100 bp
* 185957 195249: contig of 9223 bp in length
* 195250 195349: gap of 100 bp
* 195350 196868: contig of 1519 bp in length
* 196869 196968: gap of 100 bp
* 196969 198127: contig of 1159 bp in length.

FEATURES
source
1..198127
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q21"
/clone="RP11-86901"
1..58065
/note="assembly_fragment"
misc_feature
58166..92043
/note="assembly_fragment"
misc_feature
92144..111587
/note="assembly_fragment"
misc_feature
111688..129552
/note="assembly_fragment"
misc_feature
129653..149317
/note="assembly_fragment"
misc_feature
149418..164242
/note="assembly_fragment clone_end:SP6 vector_side:left"
misc_feature
164343..175076
/note="assembly_fragment clone_end:T7 vector_side:right"
misc_feature
175177..185856
/note="assembly_fragment"
misc_feature
185957..195249
/note="assembly_fragment"
misc_feature
195350..196868
/note="assembly_fragment"
misc_feature
196869..198127
/note="assembly_fragment"

ORIGIN
Query Match      85.5%; Score 18.8; DB 2; Length 198127;
Best Local Similarity 90.9%; Pred. No. 35;
```

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GACACGAGATGATGATCACC 22
Db 179594 GACACGATGATGATGATCACC 179573

RESULT 11
LOCUS CQ870293
DEFINITION Sequence 714 from Patent WO2004074320.
ACCESSION CQ870293
VERSION CQ870293.1 GI:51999890
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS Morris, D.W., Morris, D.W. and Malandro, M.S.
TITLE Novel therapeutic targets in cancer
JOURNAL Patent: WO 2004074320-A 714 02-SEP-2004;
Sagres Discovery, Inc. (US)
FEATURES
source location/Qualifiers
1..349980
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="seq 714, original 779,603 bases, replaced by-seq 714, from 0.000.001 to 0.349.980-seq 905, from 0.300.001 to 0.649.980-seq 906, from 0.600.001 to 0.779.603"

ORIGIN

Query Match 85.5%; Score 18.8; DB 6; Length 349980;
Best Local Similarity 90.9%; Pred. No. 34;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GACACGATGATGATGATCACC 22
Db 142953 GACACGATGATGATGATCACC 142974

RESULT 12
LOCUS AC125804/C
DEFINITION Rattus norvegicus clone CH230-818, WORKING DRAFT SEQUENCE, 3
ACCESSION AC125804
VERSION AC125804.4 GI:30467256
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULFILLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE
AUTHORS 1 (bases 1 to 226810)
Muzny, D., Maric, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alshoeke, S., Amin, A., Angiano, D., Ayalebechi, V., Aoyagi, A., Ayodeji, M., Baca, B., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biewald, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyte, M., Cree, A., D'Souza, L., Davila, M., L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebreyes, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,

Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Hulik, S., Hume, J., Idlebird, R., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuware, L., Louised, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Mallory, C., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenan, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwakoeleneh, O., Okunolu, G., Olarunpungu, A., Pal, S., Paks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfanckoch, C., Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Rell, B., Rell, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, M., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartbeyn, A., Sisson, I., Sitter, C. D., Smaj, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorrell, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villanar, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Woodson, H., Wolley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Unpublished
Direct Submission
2 (bases 1 to 226810)
Worley, K. C.
Direct Submission
Submitted (02-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 226810)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (09-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On May 9, 2003 this sequence version replaced gi:24940821.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GCTL
Center clone name: CH230-818
----- Summary Statistics
Assembly program: Atlas;
Consensus quality: 218425 bases at least Q40
Consensus quality: 220364 bases at least Q30
Consensus quality: 221808 bases at least Q20
Estimated insert size: 226184; sum-of-contigs estimation

Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see <http://www.hbrc.bcm.tmc.edu/docs/genbankdraftdata.html>)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of 'N', but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	224161: config of 224161 bp in length
*	224162	224261: gap of unknown length
*	224462	225658: config of 1367 bp in length
*	225629	225778: gap of unknown length
*	225729	226810: config of 1082 bp in length.

FEATURES
source
location/Qualifiers
1. .226810

Query Match	83.6%	Score	18.4	DB	2	Length	22610
Best Local Similarity	95.0%	Pred. No.	58				
Matches	19	Conservative	0	Mismatches	1	Indels	0
						Gaps	0

Qy 2 ACACGTAGATGAGTCATCA 21
Db 9048 ACAGTAGATGAGTCATCA 9025

RESULT 13	LOCUS	DEFINITION
146407 bp	146407 bp	146407 bp
DNA	DNA	DNA
linear	linear	linear
W727-MAR-2000	W727-MAR-2000	W727-MAR-2000
26bratish DNA sequence from clone CH211-279M15 in linkage group 20.	26bratish DNA sequence from clone CH211-279M15 in linkage group 20.	26bratish DNA sequence from clone CH211-279M15 in linkage group 20.
complete sequence.	complete sequence.	complete sequence.

REFERENCE 1 (pages 1 to 146407)
AUTHORS Babbage, S.
TITLE Direct Submission
JOURNAL Submitted (27-MAR-2004) Wellcome Trust Sanger Institute, Hinxton

During sequence assembly data is compared from overlapping clones where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >=

VECIOR: PIARBACZ.L.

Query March Similarity	81.8%	Score 18;	DB 5;	Length 146407;
Best Local Similarity	100.0%	Pred. No. 1e+02;		
Matches 18;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0.

QY 1 GACACGTAGATGAGTCA 18
|||
Db 7456 GACACGTAGATGAGTCA 7439

RESULT 14		
CHKOVAL5A/c		
LOCUS	5322 bp	DNA
CHKOVAL5A		linear
DEFINITION	Chicken ovalbumin gene, 5' flank.	
FEATURES		
ORIGIN		
VRTE		VRT 28-APR-1993

COMMENT
Original source text: Chicken DNA, clones pBR EcoRI [6,7].
Draft entry and computer-readable sequence for [1] kindly submitted
by J.S.Kaye, 14-OCT-1989.

Query Match	80.9%	Score 17.8	DB 5	Length 5322
Best Local Similarity	90.5%	Pred. No. 1	6e+02	
Matches 19	Conservative 0	Mismatches 2	Indels 0	Gaps 0

QY		1	GACACGTAGATGAGTCATCA	21
Db		1539	GAGACGTAGATGAGTCAGCA	1519

```
RESULT 15
AE011563
LOCUS
DEFINITION
Leptospira interrogans serovar lai str. 56601 chromosome I, section
372 of 397 of the complete sequence.
ACCESSION
AE011563
VERSION
AE011563.1
KEYWORDS
GI:24198148
ORGANISM
Leptospira interrogans serovar lai str. 56601
SOURCE
Leptospira interrogans serovar lai str. 56601
Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
REFERENCE
1 (bases 1 to 10808)
Ren, S.X., Fu, G., Jiang, X.G., Zeng, R., Miao, Y.G., Xu, H., Zhang, Y.X.,
Xiong, H., Lu, G., Lu, L.F., Jiang, H.Q., Jia, J., Tu, Y.F., Jiang, J.X.,
Gu, W.Y., Zhang, Y.Q., Cai, Z., Sheng, H.H., Yin, H.F., Zhang, Y.,
Zhu, G.F., Wan, M., Huang, H.L., Qian, Z., Wang, S.Y., Ma, W., Yao, Z.J.,
Shen, Y., Qiang, B.Q., Xia, Q.C., Guo, X.K., Danchin, A., Saint
Giron, I., Somerville, R.L., Wen, Y.M., Shi, M.H., Chen, Z., Xu, J.G.
and Zhao, G.P.
Unique physiological and pathogenic features of Leptospira
interrogans revealed by whole-genome sequencing
JOURNAL
Nature 422 (6934), 888-893 (2003)
PUBMED
12712204
AUTHORS
2 (bases 1 to 10808)
Ren, S., Fu, G., Jiang, X., Zeng, R., Xiong, H., Lu, L., Lu, G., Jiang, H.,
Ding, Y., Jia, J., Tu, Y., Gu, W., Cai, Z., Sheng, H., Yin, H., Zhang, Y.,
Zhu, G., Wang, S., Shen, Y., Qiang, B., Chen, Z., Wen, Y., Xu, J. and
Zhao, G.
Direct Submission
Submitted (12-MAR-2002) Chinese National Human Genome Center at
Shanghai, 250 Bi Bo Road, Shanghai 201203, China
Updated information will be available at our World Wide Web site
(http://www.chgc.sh.cn/lep/). Comments to the authors are
appreciated.

FEATURES
source
1. 10808
/organism="Leptospira interrogans serovar lai str. 56601"
/mol_type="genomic DNA"
/strain="56601"
/serovar="lai"
/db_xref="taxon:189518"
/chromosome="I"
/notes="Strain 56601 is maintained by the Institute of
Epidemiology and Microbiology, Chinese Academy of
Preventive Medicine"
228..722
/locus_tag="LA4087"
228..722
/locus_tag="LA4087"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AA51285.1"
/db_xref="GI:24198149"
/translacion="MFLFRLERKKKFLILILTFSSCKKDPDAVTEQEKDILIQ
ILMENESIHFKELMEKEKIPDTNOLAHAVEVLSLNGKLSQSEKQNSLKDRTDVE
EKFQFQVSSFSSENLASIKLAGAGTGFNRRYCPMVKNTWVSGTKLRNPADPMBDCG
DLVH"
701..1690
/locus_tag="LA4088"
701..1690
/locus_tag="LA4088"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AA51286.1"
/db_xref="GI:24198150"
/translacion="MMGPSTLNVVCLDKKIRCKIGVSLGPCPANCROVLLNEN
SSIRBSCLAGEKMAFAEGKAGINVASAFALPVTREIRISKRGVYELITLIFS
IIEPKKEVIFSFKLKIITIIYVNYKQTEDEBSKSLKKNASFLDIREAVYGS
SBERISFTAYCLCKAKKDPKBPESQFIVTPSPASDRLNAGLNVSEKEEYLSKDG
IYELPQIPIGIRHITSLDEDDPEITFLISTWEELVLRKQOIISSHAIILEKYSKGL
NMQSLGDLISVLELSMEIIGIIFEEKELIGRSEKMLIKELLFKHKHLKNEIT"

gene
CDS
complement(1699..2523)
/locus_tag="LA4089"
complement(1699..2523)
/locus_tag="LA4089"
/codon_start=1
/transl_table=1
/product="phenazine biosynthesis protein, PhzF family"
/protein_id="AA51287.1"
/db_xref="GI:24198151"
/translacion="MKSIIOMECSCAIFMFKLQIDSPREKVFQGNPAVCPLOQWI
SDMMOKIASENNLSERAPVVGESRRIRWFTPLBVDLCGATLAAVYHIFREGEV
IGKLNQOSISGLNFKENILYLDPPSKPEKVTPEKILNFSISKEVFKSDY
MLAYENNOUKKLEYEGRELVNTIGIATCGNDSYDPLSRFPAPSGVLEDPVATG
SAACTLIIPWSESLGKKIMKAYOTSARGGLFCEDIGERVRIGTCCVPLDGMIEI"
2752..2868
/locus_tag="LA4090"
2752..2868
/locus_tag="LA4090"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AA51288.1"
/db_xref="GI:24198152"
/translacion="MKKFRVYPTLIEFVHKIPGNSALPQKPLFRNPFII"
complement(3535..5007)
/locus_tag="LA4092"
complement(3535..5007)
/locus_tag="LA4092"
/BC_number="2.4.2.14"
/notes="Glutamine phosphoribosylpyrophosphate
amidotransferase"
/codon_start=1
/transl_table=1
/product="Amidophosphoribosyltransferase"
/protein_id="AA51289.1"
/db_xref="GI:24198153"
/translacion="MSRIIDKSRIRROAQDDPKKECAIFGIPNSSEANSFTYGLYS
MORQSSSGIVSDGELRYRAGMGLVAHIFTEKRLLEQGNALGNRRSTGASF
LMAQPLRVESHGAVSLAHNGNLVNSWELSSQLEKESIITDIDSIVIHMAKSG
ETDPLSALSKKRGAVSLVILTKTOLIVDONGRPLVMGRBRDGGIVPASETC
AFDITDYERVDGEMIVVDKGMNSYVPPASBGLCTFEYIYRPPDSSIFGES
VYKVRKNGRLPARELPVPADVVPVPSANIALGVAESGISYSGILSHYIGRT
PIEPQKIRDDGAKTKNVVNVGKVAIVDVSIMKGTYSKRIIMIRNAGAKETH
LAVSAPRTISPTGIDILPTNELIAAHTHLEIRKYLKRVDSIAYLSVESMNRVMDH
KGGFCNACTFAOYVPEFQSELGQKSLFKKEYQVEHVVY"
4950..6215
/locus_tag="LA4091"
4950..6215
/locus_tag="LA4091"
/codon_start=1
/transl_table=1
/product="Ribonuclease D"
/protein_id="AA51290.1"
/db_xref="GI:24198154"
/translacion="MVCHPGILVPLFWIYREDSRSLSLKQVAFPMQINSYIVD
TARSLQVILNLSQADSIVDTSESSGYTTFKVCGLQISAKGNVITDPLKQNLSE
IGNIFEDKTIKIFHSAIDIDIKALKKQFGQFOQNIADTGSRLDHEOYSTLYVDY
YHKIUSKRDQKSWERPLEKQLOQAADTYVLETIWEMKEELIKMLYEEALSE
FEKIASSEPGSEGNISMDKFPETILEYSDERFIYDTLVFRDQSKLKNAPFAVSE
NEKVVLLMKARDMKRLTEIVGKKDAETLQIYANPSGPPIDQSKLPKGEVNTDE
GERFKRLIRWETIMSIIRMSHOMBPENKTIQALISQNPRTIDELKMLFSEKVIH
YGPILALANVPESWIKGLIPINKKE"
6247..6918
/locus_tag="LA4093"
6247..6918
/locus_tag="LA4093"
/codon_start=1
/transl_table=1
/product="MutT/nudix family protein"
/protein_id="AA51291.1"
/db_xref="GI:24198155"
/
```

```
OY      1 GACACGTAAGATGAGTCATCA 21
          ||| |||| |
DB      5999 GATACGAAGAATGAGTCATCA 6019
```

Search completed: March 1, 2005, 00:29:25
Job time : 648.714 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 28, 2005, 23:35:01 : Search time 2081.3 Seconds
(without alignments)
548.661 Million cell updates/sec

Title: US-10-086-062-4

Perfect score: 30
Sequence: 1 ctggacccctcgcagcgcagagctcgcgt 30

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapect 1.0

Searched: 34239544 seqs, 19032134700 residues
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:
1: gb_ests1:
2: gb_ests2:
3: gb_ests3:
4: gb_ests4:
5: gb_ests5:
6: gb_ests6:
7: gb_ests7:
8: gb_ests8:
9: gb_ests9:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20.6	68.7	1091	5	BM905373 AGENCOURT
2	20.4	68.0	804	4	BG430455 602502021
3	20.2	67.3	877	7	CK873283 AGENCOURT
4	20.2	67.3	1218	4	BG474756 602517583
5	20	66.7	838	7	CN127674 RHOM 24
6	19.6	65.3	634	9	CE429361 CIGR-GB8-
7	19.4	64.7	123	2	BF889056 MR4-TN011
8	19.4	64.7	168	9	BF889056 MR4-TN011
9	19.4	64.7	240	4	BF889056 MR4-TN011
10	19.4	64.7	245	2	BF889056 MR4-TN011
11	19.4	64.7	254	2	BF889056 MR4-TN011
12	19.4	64.7	469	4	BF889056 MR4-TN011
13	19.4	64.7	511	4	BF889056 MR4-TN011
14	19.4	64.7	541	4	BF889056 MR4-TN011
15	19.4	64.7	564	4	BF889056 MR4-TN011
16	19.4	64.7	588	1	BF889056 MR4-TN011
17	19.4	64.7	688	1	BF889056 MR4-TN011
18	19.4	64.7	893	7	BF889056 MR4-TN011
19	19.4	64.7	893	7	BF889056 MR4-TN011
20	19.4	64.7	893	7	BF889056 MR4-TN011
21	19.4	64.7	893	7	BF889056 MR4-TN011
22	19.4	64.7	893	7	BF889056 MR4-TN011
23	19.4	64.7	893	7	BF889056 MR4-TN011
24	19.4	64.7	893	7	BF889056 MR4-TN011

25	19	63.3	504	5	BX087591
26	19	63.3	517	7	CK518339
27	19	63.3	522	8	A0229833 HS 2027 A
28	19	63.3	525	7	CK885179 SGP165576
29	19	63.3	526	2	BE684663
30	19	63.3	535	5	BX301244
31	19	63.3	544	7	CK898692 SGP162270
32	19	63.3	552	5	BX317355
33	19	63.3	558	5	BX888009
34	19	63.3	586	6	A2806755
35	19	63.3	598	8	CA343648
36	19	63.3	625	4	BM172236
37	19	63.3	630	5	BX877903
38	19	63.3	633	5	BX317619
39	19	63.3	680	5	BX319301
40	19	63.3	709	2	AM743305
41	19	63.3	730	5	BX861731
42	19	63.3	751	5	BX297481
43	19	63.3	752	6	CA353623
44	19	63.3	756	2	BF338402
45	19	63.3	763	5	BX873117

ALIGNMENTS

RESULT 1
BM905373
LOCUS
DEFINITION
AGENCOURT 6698806 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5558278
5', mRNA sequence.
BM905373
BM905373.1 GI:19355752
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 1091)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaaba-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM1280 row: p column: 23
High quality sequence stop: 573.
Location/Qualifiers
1..1091
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5558278"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_id="NIH MGC 72"
/note="Organ: skin; Vector: pCMV-SORT6; Site: 1; NCI;
Site: 2; Salt; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."

ORIGIN
Query Match 68.7% Score 20.6; DB 5; Length 1091;
Best Local Similarity 85.2% Pred. No. 3.1e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
1 CTGACCCCTCTGACGACGAGATTCC 27

Db 900 CTGAGACCCCTCTCGACTCGGAGATCCC 926

|||||

RESULT 2
BG430455/c 804 bp mRNA 1linear EST 14-MAR-2001
LOCUS 602502021F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4615756 5',
DEFINITION mRNA sequence.
ACCESSION BG430455
VERSION BG430455.1 GI:13337063
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE NIH-MGC http://mgs.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CLOUTCH Laboratories, Inc.
CDNA Library Preparation: CLOUTCH Laboratories, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNCM1369 row: 1 column: 05
High quality sequence stop: 742.
Location/Qualifiers

FEATURES
source
1. 804
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4615756"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_75"
/note="Organ: Kidney; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggcgccgctggc); Site 2: SfiI (ggcgccgctggc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCGCATATGATG-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGGAGCGGCGGAGATG-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.65
kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 68.0%; Score 20.4; DB 4; Length 804;
Best Local Similarity 80.0%; Pred. No. 3.7e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGAGACCCCTCTCGACTCGGAGATCCC 30
Db 559 CTGAGACCCCTCTCGACTCGGAGATCCC 530

|||||

RESULT 3
CK873283 877 bp mRNA 1linear EST 09-MAR-2004
LOCUS AGENCOURT 19436603 NIH_ZGC_16 Danio rerio cDNA clone IMAGE:7215126
DEFINITION 5', mRNA sequence.
ACCESSION CK873283
VERSION CK873283.1 GI:45289381
KEYWORDS EST.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
JOURNAL Cypriniformes; Cyprinidae; Danio.
COMMENT

REFERENCE 1 (bases 1 to 877)
AUTHORS NIH-MGC http://mgs.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgabbs-remail.nih.gov
Tissue Procurement: Will Talbot
CDNA Library Preparation: Dr. Yutaka Suzuki and Dr. Sumio Sugano
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM15111 row: h column: 04
High quality sequence start: 13
High quality sequence stop: 685.
Location/Qualifiers

FEATURES
source
1. 877
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:79955"
/clone="IMAGE:7215126"
/issue_type="13 pooled, includes stomach, intestine,
liver and pancreas"
/lab_host="DH10B Tona"
/clone_lib="NIH_ZGC 16"
/note="Organ: gut; Vector: pME18S-FL3; Site 1: DraIII;
Site 2: DraIII; 1st strand cDNA was primed with an
oligo(dT) primer
[GGGCTGAAGAGCGCTATGTCGCTTTTCTTTTCTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[GGCCUACUGG], digested and directionally cloned into
distinct DraIII sites of the pME18S-FL3. Library was size
selected for 1.0 kb with a average insert size of ~1.2kb.
Library constructed by Yutaka Suzuki (University of Tokyo
Institute of Medical Science). Custom primers recommended
for sequencing: 5' end primer 5'-GGATGTCCTTTACTCTTA-3'
and 3' end primer 5'-CGACCTGACGCTGACACCA-3'. Note: This
is a Zebrafish Gene Collection (ZGC) library"

ORIGIN

Query Match 67.3%; Score 20.2; DB 7; Length 877;
Best Local Similarity 88.0%; Pred. No. 4.5e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TGAACCCCTCTCGACTCGGAGATTC 26
Db 817 TGAACCCCTCTCGACTCGGAGATTC 793

|||||

RESULT 4
BG474756 1218 bp mRNA 1linear EST 21-MAR-2001
LOCUS 602517583F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:4649109 5',
DEFINITION mRNA sequence.
ACCESSION BG474756
VERSION BG474756.1 GI:13407020
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE NIH-MGC http://mgs.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: InCyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1CM1428 row: f column: 22
High quality sequence stop: 1.
Location/Qualifiers

FEATURES

source

1..1218
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4649109"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_16"
/note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 67.3%; Score 20.2; DB 4; Length 1218;
Best Local Similarity 88.0%; Pred. No. 4.6e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 TGAGACCCCTCTGACTCGAGAGTTC 26
|||||
Db 769 TGAACCTCTGACTCGAGAGCTC 793

RESULT 5
LOCUS CN127674/c 838 bp mRNA linear EST 01-APR-2004
DEFINITION RHOH1_24_A01.g3_A002 Acid- and alkaline-treated roots Sorghum
bicolor CDNA clone RHOH1_24_A01_A002 5', mRNA sequence.
CN127674
CN127674.1 GI:45953877
EST.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE

AUTHORS

TITLE
JOURNAL
COMMENT

1 (bases 1 to 838)
Cordonnier-Pratt M.-M., Suzuki Y., Sugano S., Klein R.R., Liang C., Sun F., Sullivan R., Lim S., Eastman A. and Pratt L.H.
An EST database from Sorghum: acid- and alkaline-treated roots
Unpublished (2003)
Other ESTs: RHOH1_24_A01.b1_A002
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; plant material and RNA prepared at Texas A & M University; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.
Seq primer: Sugs (CTTCTGCTTAAGAGCTCGC).

FEATURES

source

1..838
/organism="Sorghum bicolor"
/mol_type="mRNA"

/cultivar="BTx623"
/db_xref="taxon:4558"
/clone="RHOH1_24_A01_A002"
/lab_host="DH10B-T1 phage-resistant E. coli"
/clone_lib="Acid- and alkaline-treated roots"
/note="Organ: Root; Vector: pME18S-FL3; Site_1: XhoI; Site_2: XhoI; The library was prepared from polyA+ RNA from 8-day-old roots harvested from BTx623 sorghum seedlings grown in hydroponic culture. HCl was added to a pH of 3.0 to some seedlings, KOH to a pH of 9.0 for others. Roots were harvested 3, 12 and 27 hr after addition and pooled for RNA preparation. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5'-prime DraIII site is CACTGTGG, 3'-prime DraIII site is CACCATGG). XhoI excises the cDNA insert."

ORIGIN

Query Match 66.7%; Score 20; DB 7; Length 838;
Best Local Similarity 82.1%; Pred. No. 5.5e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CTGACCCCTCTGACTCGAGAGTTCG 28
|||||
Db 526 CTGACCCCAAGCATCGAGAGTTCG 499

RESULT 6
LOCUS CE429361/c 634 bp DNA linear GSS 27-SEP-2003
DEFINITION tigr-gss-dog-17000362924386 Dog library Canis familiaris genomic,
genomic survey sequence.
CE429361
CE429361.1 GI:36701435
GSS.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Canis familiaris (dog)
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euteria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 634)
Kirknes E.F., Bafna V., Halpern A.L., Levy S., Remington K.,
Rusch D.B., Delcher A.L., Pop M., Wang W., Fraser C.M. and
Venter J.C.
The dog genome: survey sequencing and comparative analysis
Science 301 (5641), 1898-1903 (2003)

TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

Contact: Kirknes EF
The Institute for Genomic Research
Department of Bukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
Location/Qualifiers

FEATURES

source

1..634
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site_1: BclII; Libraries were prepared from peripheral blood"

ORIGIN

Query Match 65.3%; Score 19.6; DB 9; Length 634;
Best Local Similarity 84.6%; Pred. No. 8.2e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 TGAGACCCCTCTGACTCGAGAGTTC 27
|||||
Db 368 TGAACCTCTGAGCTAGAGAGCTCC 343

RESULT 7
BF889056/c 123 bp mRNA linear EST 18-JAN-2001
LOCUS MR4-TN0116-271100-202-e06 TN0116 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF889056
ACCESSION BF889056.1 GI:12280302
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 123)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.G.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
JOURNAL MEDLINE
PUBMED 20202663
COMMENT 10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=MR4cl2-MR4-TN0116-271100-202-e06&cl=2000-11-27&cl=1)
Seq primer: puc 18 forward
High quality sequence start: 68
High quality sequence stop: 123.
Location/Qualifiers
1..123
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="TN0116"
/note="Organ: testis normal; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORFESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
ORIGIN
Query Match 64.7%; Score 19.4; DB 2; Length 123;
Best Local Similarity 79.3%; Pred. No. 9.6e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 CTGACCCCTCTGACTCGAGAGTTCGCC 29
41 CTGACCCCTCTGACTCGAGAGTTCGCC 13
Db
RESULT 8
LBAF041G06 168 bp DNA linear GSS 23-JUN-2003
LOCUS LBAF041G06
DEFINITION Leishmania braziliensis GSS, clone LBAF041G06, genomic survey
ACCESSION EX345620
VERSION EX345620.1 GI:32170214
KEYWORDS GSS; genomic survey sequence.

SOURCE
ORGANISM Leishmania braziliensis
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania; Leishmania braziliensis species complex.
REFERENCE 1
Laurentino, E.C., Ruiz, J.C. and Cruz, A.K.
GSS analysis of the Leishmania braziliensis genome
Unpublished
2 (bases 1 to 168)
REFERENCE
AUTHORS Cruz, A.K.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-2003) Cruz A.K., University of Sao Paulo, Department of Molecular and Cell Biology, FMBP, Avenida Bandeira 3900, Ribeirao Preto, SP, 14049-900, BRAZIL
Clone requests: akcruz@fmbp.usp.br.
COMMENT
FEATURES
SOURCE
1..168
/organism="Leishmania braziliensis"
/mol_type="genomic DNA"
/strain="MHOM/BR/75/M2904"
/db_xref="taxon:5660"
/clone="LBAF041G06"
ORIGIN
Query Match 64.7%; Score 19.4; DB 9; Length 168;
Best Local Similarity 79.3%; Pred. No. 9.6e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 CTGACCCCTCTGACTCGAGAGTTCGCC 29
43 CTGACCCCTCTGACTCGAGAGTTCGCC 71
Db
RESULT 9
BI055318/c 240 bp mRNA linear EST 15-JUN-2001
LOCUS RCO-GN0233-060201-031-f07 GN0233 Homo sapiens cDNA, mRNA sequence.
DEFINITION BI055318
ACCESSION BI055318.1 GI:14462848
VERSION EST.
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 240)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.G.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
JOURNAL MEDLINE
PUBMED 20202663
COMMENT 10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=RCOcl2-RCO-GN0233-060201-031-f07&cl=2001-02-06&cl=1)
Seq primer: puc 18 forward
High quality sequence start: 69
High quality sequence stop: 156.
Location/Qualifiers
1..240

ORIGIN

Qy 1 CTGGACCCCTCTCGACTCGAGA GTTCCGC 23

Ddb 55 CTGGAGCCTGCTCGACTCTAGAGATCCTC 27

REFERENCE AUTHORS

PUBMED
COMMENT

FEATURES
SOURCE

```

1. .245
/organisms="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="P02000"
/notes="Organ: prostate tumor; Vector: puc18; Site_1: Small
Site_2: Small; A mini-library was made by cloning products
derived from ORSITES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of

```

ORIGIN

Qy 1 CTGACCCCTCTCGACTCGAGATTCCGC 25
||| ||| ||| ||| ||| ||| ||| |||
Db 59 CTGGAGCCTGTCTGACTCTAGATTCAC 311

SOURCE ORGANISM

REFERENCE AUTHORS

PUBMED
COMMENT

FEATURES.
SOURCE

ORIGIN

Query Match	64.7%	Score 19.4	DB 2	Length 254
Best Local Similarity	79.3%	Pred. No.	9.7e+02	
Matches 23	Conservative 0	Mismatches 6	Indels 0	Gaps 0
Qy	1	CTGGACCCCTTCGACTCGAGATTTCGC	29	

1 CTGGACCCCTCTGACTCGAGATTCCGC 25

Db 29 CTGAGCTGCTGCTGACTTAGAGATCCAC 1

RESULT 12

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

FEATURES

source

1. 469
/organism="Apis mellifera"
/mol_type="mRNA"
/strain="mixed strains of European bees, predominantly
A.m. ligustica"
/db_xref="taxon:7460"
/clone="BB160023B10C10"
/sex="female"
/tissue_type="brain"
/dev_stage="adult worker honey bee"
/lab_host="DH10B"
/clone_lib="Bee Brain Normalized Library, BB16"
/note="Organ: brain; Vector: pTV3-Pac; Site: 1: EcoRI;
Site 2: NotI. The BB16 library was contributed by the
Soares laboratory and it was constructed and normalized
as described by Bonaldo, M.F., Lennon, G. and Soares,
M.B. (1996), Genome Research 6(9): 791-806. RNA was
prepared from dissected brains of adult worker bees of
various ages and various behavioral groups."

ORIGIN

Query Match

Best Local Similarity 64.7%; Score 19.4; DB 4; Length 469;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 7 CCCTTCGACTCGAGAGTTCC 27
|||||
Db 273 CCCTTCGACTCGAGAGTTCC 293

RESULT 13

BB1514171

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

FEATURES

source

1. 511
/organism="Apis mellifera"
/mol_type="mRNA"
/strain="mixed strains of European bees, predominantly
A.m. ligustica"
/db_xref="taxon:7460"
/clone="BB160014B10C05"
/sex="female"
/tissue_type="brain"
/dev_stage="adult worker honey bee"
/lab_host="DH10B"
/clone_lib="Bee Brain Normalized Library, BB16"
/note="Organ: brain; Vector: pTV3-Pac; Site: 1: EcoRI;
Site 2: NotI. The BB16 library was contributed by the
Soares laboratory and it was constructed and normalized
as described by Bonaldo, M.F., Lennon, G. and Soares,
M.B. (1996), Genome Research 6(9): 791-806. RNA was
prepared from dissected brains of adult worker bees of
various ages and various behavioral groups."

ORIGIN

Query Match

Best Local Similarity 64.7%; Score 19.4; DB 4; Length 511;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 7 CCCTTCGACTCGAGAGTTCC 27
|||||
Db 274 CCCTTCGACTCGAGAGTTCC 294

RESULT 14

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

ACCESSION BI513884
 VERSION BI513884.1 GI:15364258
 KEYWORDS
 SOURCE Apis mellifera (honey bee)
 ORGANISM Apis mellifera
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea; Apidae; Apis.
 REFERENCE 1 (bases 1 to 541)
 Whitfield,C.W., Band,M.R., Bonaldo,M.F., Kumar,C.G., Liu,L., Pardinas,J., Robertson,H.M., Soares,B. and Robinson,G.E.
 Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee
 JOURNAL Genome Res. 12 (4), 555-566 (2002)
 MEDLINE 21929762
 PUBMED 11932240
 COMMENT Contact: Gene E. Robinson
 Department of Entomology
 University of Illinois
 505 S. Goodwin Ave., Urbana, IL 61801, USA
 Tel: 217 265 0309
 Fax: 217 244 3499
 Email: generobi@life.uiuc.edu
 This research was funded by the University of Illinois Critical Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation Award in Functional Genomics to G.E. Robinson and an NSF Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
 PCR Primers
 FORWARD: TAATAGACTCACTATAGG
 BACKWARD: ATTAACCTCTACTAAG
 Plate: BB160014A11 row: F column: 07
 Seq primer: AGCGGATACAAATTCACACAGA
 High quality sequence stop: 541.
 Location/Qualifiers
 1..541
 /organism="Apis mellifera"
 /mol_type="mRNA"
 /strain="mixed strains of European bees, predominantly A.m. ligustica"
 /db_xref="taxon:7460"
 /clone="BB160014A11F07"
 /sex="female"
 /tissue_type="brain"
 /dev_stage="adult worker honey bee"
 /lab_host="DH10B"
 /clone_1lb="Bee Brain Normalized Library, BB16"
 /note="Organ: Brain; Vector: pRTT3-Pac; Site: 1: Ecot1; Site 2: Not1; The BB16 library was contributed by the Soares laboratory and it was constructed and normalized as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. RNA was prepared from dissected brains of adult worker bees of various ages and various behavioral groups."

ORIGIN
 Query Match 64.7%; Score 19.4; DB 4; Length 541;
 Best Local Similarity 95.2%; Pred. No. 9.9e+02;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 7 CCCTCTCGACTCGAGGTTCC 27
 |||||
 Db 182 CCCTCTCGACTCGAGGTTCC 202

RESULT 15
 BI513971
 LOCUS BB160014A11P07.5 Bee Brain Normalized Library, BB16 Apis mellifera
 DEFINITION CDNA clone BB160014A11P07.5, mRNA sequence.
 ACCESSION BI513971
 VERSION BI513971.1 GI:15364345
 KEYWORDS
 SOURCE Apis mellifera (honey bee)

ORGANISM Apis mellifera
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea; Apidae; Apis.
 REFERENCE 1 (bases 1 to 564)
 Whitfield,C.W., Band,M.R., Bonaldo,M.F., Kumar,C.G., Liu,L., Pardinas,J., Robertson,H.M., Soares,B. and Robinson,G.E.
 Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee
 JOURNAL Genome Res. 12 (4), 555-566 (2002)
 MEDLINE 21929762
 PUBMED 11932240
 COMMENT Contact: Gene E. Robinson
 Department of Entomology
 University of Illinois
 505 S. Goodwin Ave., Urbana, IL 61801, USA
 Tel: 217 265 0309
 Fax: 217 244 3499
 Email: generobi@life.uiuc.edu
 This research was funded by the University of Illinois Critical Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation Award in Functional Genomics to G.E. Robinson and an NSF Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
 PCR Primers
 FORWARD: TAATAGACTCACTATAGG
 BACKWARD: ATTAACCTCTACTAAG
 Plate: BB160014A11 row: F column: 07
 Seq primer: AGCGGATACAAATTCACACAGA
 High quality sequence stop: 564.
 Location/Qualifiers
 1..564
 /organism="Apis mellifera"
 /mol_type="mRNA"
 /strain="mixed strains of European bees, predominantly A.m. ligustica"
 /db_xref="taxon:7460"
 /clone="BB160014A11F07"
 /sex="female"
 /tissue_type="brain"
 /dev_stage="adult worker honey bee"
 /lab_host="DH10B"
 /clone_1lb="Bee Brain Normalized Library, BB16"
 /note="Organ: Brain; Vector: pRTT3-Pac; Site: 1: Ecot1; Site 2: Not1; The BB16 library was contributed by the Soares laboratory and it was constructed and normalized as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. RNA was prepared from dissected brains of adult worker bees of various ages and various behavioral groups."

ORIGIN
 Query Match 64.7%; Score 19.4; DB 4; Length 564;
 Best Local Similarity 95.2%; Pred. No. 9.9e+02;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 7 CCCTCTCGACTCGAGGTTCC 27
 |||||
 Db 182 CCCTCTCGACTCGAGGTTCC 202

Search completed: March 1, 2005, 01:58:25
 Job time : 2088.63 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 28, 2005, 23:26:41 ; Search time 261.818 Seconds
(without alignments)
678.304 Million cell updates/sec

Title: US-10-086-062-4

Perfect score: 30
Sequence: 1 ctgagccctctcgtactcagagagtcgcgt 30

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneeqn1980s:*
2: geneeqn1990s:*
3: geneeqn2000s:*
4: geneeqn2001as:*
5: geneeqn2001bs:*
6: geneeqn2002as:*
7: geneeqn2002bs:*
8: geneeqn2003as:*
9: geneeqn2003bs:*
10: geneeqn2003cs:*
11: geneeqn2003ds:*
12: geneeqn2004as:*
13: geneeqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	30	100.0	30	AAD24069	Aad24069 Maize eng
2	19.2	64.0	1500	5 AAS11024	Aas11024 Vibrio ch
3	19	63.3	121124	12 ADG97107	Adg97107 Mouse can
4	18.8	62.7	2461	10 ABT41766	Abt41766 Toxicity
5	18.8	62.7	23107	9 ADA02762	Ada02762 Human RUN
6	18.8	62.7	23107	10 ADB72500	Adb72500 Human Run
7	18.8	62.7	23107	10 ADC85242	Adc85242 Human Run
8	18.8	62.7	23107	12 ADM74357	Adm74357 Human car
9	18.6	62.0	3210	2 AAZ27624	Aaz27624 Plasmid S
10	18.6	62.0	3459	2 AAZ27623	Aaz27623 Plasmid S
11	18.4	61.3	319608	3 AAHS1601	Aahs1601 Human chr
12	18.4	61.3	319608	5 AAS09301	Aas09301 Human sch
13	18.2	60.7	340449	8 AAL52198	Aal52198 Human sec
14	18	60.0	42325	10 ADB74382	Adb74382 Mycobacte
15	18	60.0	110000	11 ADM27081_10	Adm27081_10 Contination (11 O
16	17.8	59.3	165	10 ADD49400	Add49400 Human
17	17.8	59.3	254	12 ADG00322	Adg00322 Nicotiana
18	17.8	59.3	369	11 ABD11207	Abd11207 Pseudomon
19	17.8	59.3	426	10 ADD49385	Add49385 Human lun
20	17.8	59.3	449	10 ADD49343	Add49343 Human lun

21	17.8	59.3	449	10	ADD49294	Add49294 Human lun
22	17.8	59.3	986	6	ABQ46720	Abq46720 Oligonuc1
23	17.8	59.3	986	6	ABQ46721	Abq46721 Oligonuc1
24	17.8	59.3	1253	12	ADL12873	Adl12873 Human ste
25	17.8	59.3	1603	4	AAK77077	Aak77077 Human lmm
26	17.8	59.3	1731	11	ACN91401	Acn91401 Breast ca
27	17.8	59.3	2780	13	ADR25834	Adr25834 Breast ca
28	17.8	59.3	152141	8	ACA64961	Ac64961 Human BCR
29	17.8	59.3	15395	13	ADP33534	Adp33534 Murine ca
30	17.6	58.7	246	12	ADQ52469	Adq52469 Human met
31	17.6	58.7	8119	3	AAZ35392	Aaz35392 Maize sca
32	17.6	58.7	49243	4	ABL03188	Ab103188 Drosoph11
33	17.4	58.0	65	6	ABN51822	Abn51822 Mouse spl
34	17.4	58.0	300	2	AAZ13036	Aaz13036 Human gen
35	17.4	58.0	300	2	AAZ98464	Aaz98464 Human gen
36	17.4	58.0	411	4	AA185652	Aa185652 Human pol
37	17.4	58.0	460	9	ACH41661	Ach41661 Human foe
38	17.4	58.0	460	9	ACH74968	Ach74968 Human gen
39	17.4	58.0	699	2	AAZ15926	Aaz15926 Human gen
40	17.4	58.0	729	12	ADO63441	Ado63441 Transcrip
41	17.4	58.0	778	6	ABK30430	Abk30430 Novel S
42	17.4	58.0	936	13	ADR92877	Adr92877 Bacterial
43	17.4	58.0	949	13	ADS50577	Aas50577 Bacterial
44	17.4	58.0	1210	4	AA641038	Aa641038 CDNA enco
45	17.4	58.0	1210	4	AA634819	Aa634819 CDNA enco

ALIGNMENTS

RESULT 1	
AAD24069	
ID AAD24069	standard; DNA; 30 BP.
XX	
AC AAD24069;	
XX	
DT 09-APR-2002	(first entry)
XX	
DE Maize engineered Ubi-1 promoter heat shock element #3.	
XX	
KW Gene expression; maize; ubiquitin promoter; Ubi-1; HSE;	
KW heat shock element; agronomic gene; ds.	
XX	
OS Zea mays.	
XX	
FH Key	Location/Qualifiers
FT misc_feature	1..15
FT	/tag= a
FT	/note= "5' heat shock element"
FT	16..30
FT	/tag= b
FT	/note= "3' heat shock element"
XX	
PN WO200194394-A2.	
XX	
PD 13-DEC-2001.	
XX	
PF 08-JUN-2001; 2001WO-US018689.	
XX	
PR 09-JUN-2000; 2000US-00530558.	
XX	
PA (PROD-) PRODIGENE INC.	
XX	
PI Jilka JM, Hood EE, Howard JA;	
XX	
DR WPI; 2002-122117/16.	
XX	
PT New promoter sequences for causing expression of a structural gene	
PT especially agronomic gene or open reading frame in a plant cell,	
PT comprises engineered versions of the maize ubiquitin promoter.	
XX	
PS Claim 6; Page 54; 68pp; English.	
XX	

CC The invention relates to a promoter sequence capable of directing
CC expression of a nucleotide sequence in a plant cell, comprising maize
CC ubiquitin (ubi-1) promoter sequence with a modification so that it does
CC not include two overlapping heat shock elements (HSE) or it directs
CC expression to increase the endosperm/embryo expression ratio of the
CC protein when compared to the ratio from a wild-type ubiquitin promoter.
CC The modified ubi-1 promoter comprises a deletion of 3', 5' or both HSEs,
CC two non-overlapping/adjacent HSEs, replacement of HSEs with a trimer of a
CC seed specific element from the promoter of pea lectin gene *Pel*, or
CC insertion of a transcription factor binding site in the HSE region. An
CC expression construct comprising modified ubi-1 promoter is useful for
CC causing expression of a structural gene (agronomic genes) or open reading
CC frame in a plant cell. The modified ubi-1 promoter increases expression
CC levels beyond those observed with native ubiquitin promoter. The present
CC sequence is maize engineered ubi-1 promoter with heat shock elements
CC adjacently placed. Note: The present sequence is also shown in claim 26,
CC page-56 of the specification. However, this sequence has an additional
CC nucleotide at the 3' end
XX

SO Sequence 30 BP; 4 A; 12 C; 7 G; 7 T; 0 U; 0 Other;

Query Match 100.0%; Score 30; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGGACCCCTCTCGACTCGAGATTCCGCT 30
DB 1 CTGGACCCCTCTCGACTCGAGATTCCGCT 30

RESULT 2

AAS11024/C
ID AAS11024 standard; DNA; 1500 BP.

XX AAS11024;
AC 11-SEP-2003 (revised)
DT 24-OCT-2001 (first entry)
XX
DE Vibrio cholera 16S ribosomal RNA gene.
XX
KM Antisense; bacterial 16S ribosomal RNA; rRNA; bacterial infection; human;
XX food grain supplement; livestock; poultry; therapeutic; ds.
XX
OS Vibrio cholerae.
XX
PN WO200142457-A2.
XX
PD 14-JUN-2001.
XX
PF 29-NOV-2000; 2000WO-US042391.
XX
PR 29-NOV-1999; 99US-0168150P.
XX
PA (AVTB-) AVT BIOPHARMA INC.
XX
PI Iversen PL;
XX
DR WPI; 2001-457295/49.
XX
PT Antibacterial compound, useful for treating bacterial infections and as
PT livestock and poultry food supplement, comprises antisense
PT oligonucleotides complementary to bacterial 16S and 23S rRNA.
XX
PS Disclosure; Page; 62pp; English.
XX

CC AAS11021-AAS11034 represent the coding sequences of bacterial 16S
CC ribosomal RNA (rRNA) genes. The sequences were used to design anti-
CC bacterial compounds comprising substantially uncharged antisense
CC oligomers containing 8-40 nucleotide subunits, including a targeting
CC nucleic acid sequence at least 10 nucleotides in length which is
CC complementary to a bacterial 16S or 23S rRNA nucleic acid sequence. The
CC antisense oligomers are used for treating a bacterial infection in a

CC human or a mammalian animal produced by *Escherichia coli*, *Salmonella*
CC typhimurium, *Pseudomonas aeruginosa*, *Vibrio cholera*, *Neisseria*
CC gonorrhoea, *Helicobacter pylori*, *Bartonella henselae*, *Haemophilus*
CC influenza, *Shigella dysenteriae*, *Staphylococcus aureus*, *Mycobacterium*
CC tuberculosis, *Streptococcus pneumoniae*, *Treponema pallidum* and *Chlamydia*
CC trachomatis. The antibacterial compound may be used as a food grain
CC supplement in livestock and poultry food composition. Note: The present
CC sequence is not shown in the specification but has been accessed from
CC Genbank using the appropriate accession number given in the
CC specification. (Updated on 11-SEP-2003 to standardise OS field)
XX

SO Sequence 1500 BP; 376 A; 326 C; 482 G; 312 T; 0 U; 4 Other;

Query Match 64.0%; Score 19.2; DB 5; Length 1500;
Best Local Similarity 75.0%; Pred. No. 1.1e+02;
Matches 21; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 CTGGACCCCTCTCGACTCGAGATTCCG 28
DB 212 CTGGACCCCTCTCGACTCGAGATTCCG 185

RESULT 3

ADQ97107/C
ID ADQ97107 standard; DNA; 121124 BP.

XX ADQ97107;
AC ADQ97107;
DT 07-OCT-2004 (first entry)
XX
DE Mouse cancer associated sequence MD08-002, SEQ ID 83.
XX
KM Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Mouse; ds.
XX
OS Mus musculus.
XX
PN WO2004060304-A2.
XX
PD 22-JUL-2004.
XX
PF 22-DEC-2003; 2003WO-US041389.
XX
PR 27-DEC-2002; 2002US-00330773.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Malandro MS;
XX
DR WPI; 2004-543781/52.
XX
DT New isolated cancer associated nucleic acids comprising at least 10
DT contiguous nucleotides, useful for diagnosing, preventing and/or treating
DT cancers such as leukemia and lymphoma.
XX
PS Claim 1; SEQ ID NO 83; 199pp; English.
XX
XX The present invention relates to cancer associated sequences (ADQ97025-
CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX

SO Sequence 121124 BP; 32972 A; 25314 C; 26641 G; 35451 T; 0 U; 746 Other;

Query Match 63.3%; Score 19; DB 12; Length 121124;
Best Local Similarity 81.5%; Pred. No. 1.9e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 GACCCCTCTCGACTCGAGATTCCGCT 30
DB 88764 GACCCCTCTCGACTCGAGATTCCGCT 88738

RESULT 4
 ABT41766/C
 ID ABT41766 standard; DNA; 2481 BP.
 XX
 AC ABT41766;
 XX
 DT 26-JUN-2003 (first entry)
 XX
 DE Toxicity modelling related rat gene SEQ ID NO 1468.
 XX
 KW Toxic effect; gene expression profile; renal toxicity; toxicity marker;
 KW database; drug screening; toxicity assay; rat; ds.
 OS Rattus norvegicus.
 XX
 PN WO200295000-A2.
 XX
 PD 28-NOV-2002.
 XX
 PF 22-MAY-2002; 2002WO-US016173.
 XX
 PR 22-MAY-2001; 2001US-0292335P.
 PR 13-JUN-2001; 2001US-0297523P.
 PR 19-JUN-2001; 2001US-0296925P.
 PR 10-JUL-2001; 2001US-0303807P.
 PR 10-JUL-2001; 2001US-0303808P.
 PR 10-JUL-2001; 2001US-0303810P.
 PR 28-AUG-2001; 2001US-0315047P.
 PR 27-SEP-2001; 2001US-0324928P.
 PR 22-OCT-2001; 2001US-0330462P.
 PR 01-NOV-2001; 2001US-0330867P.
 PR 21-NOV-2001; 2001US-0331805P.
 PR 06-DEC-2001; 2001US-0336144P.
 PR 19-DEC-2001; 2001US-0340873P.
 PR 21-FEB-2002; 2002US-0357842P.
 PR 21-FEB-2002; 2002US-0357843P.
 PR 21-FEB-2002; 2002US-0357844P.
 PR 15-MAR-2002; 2002US-0364134P.
 PR 08-APR-2002; 2002US-0370144P.
 PR 08-APR-2002; 2002US-0370206P.
 PR 17-APR-2002; 2002US-0372794P.
 PR 21-APR-2002; 2002US-0371679P.
 XX
 PA (GENE-) GENE LOGIC INC.
 XX
 PI Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
 XX
 DR WPI; 2003-148464/14.
 XX
 PT Predicting at least one toxic effect of a compound, useful for toxicity
 PT modelling, comprises preparing a gene expression profile of a tissue or
 PT cell sample exposed to the compound, and comparing the gene expression
 PT profile to a database.
 XX
 PS Example 4; Page; 446pp; English.
 XX
 CC The invention relates to a novel method of predicting at least one toxic
 CC effect of a compound. The method comprises a gene expression profile of a
 CC tissue or cell sample exposed to the compound, and comparing the gene
 CC expression profile to a database comprising at least part of the data or
 CC information given in the specification. The methods are useful for
 CC predicting at least one toxic effect of a compound, predicting the renal
 CC progression of a toxic effect of a compound, predicting the renal
 CC toxicity of a compound, or identifying toxicity markers in tissues or
 CC cells exposed to known renal toxin. The genes are useful as toxicity
 CC markers in drug screening and toxicity assays, in monitoring disease or
 CC physiological states, or disease progression. This polynucleotide
 CC represents a rat DNA sequence relating to the toxic effect database
 CC described in the specification. NOTE: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from the World Intellectual Property

CC Organization
 XX
 SQ Sequence 2481 BP; 539 A; 653 C; 555 G; 674 T; 0 U; 0 Other;
 XX
 Query Match 62.7%; Score 18.8; DB 10; Length 2481;
 Best Local Similarity 76.7%; Pred. No. 1.7e+02;
 Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 QY 1 CTGGACCCCTCTCGAGGAGTCCGCT 30
 DB 1077 CTGGACCCCTCTCGTACGAGGTCCAGGT 1048
 XX
 RESULT 5
 ADA02762
 ID ADA02762 standard; DNA; 23107 BP.
 XX
 AC ADA02762;
 XX
 DT 06-NOV-2003 (first entry)
 XX
 DE Human RUNX3 carcinoma associated gene, SEQ ID NO:1280.
 XX
 KW Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
 KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO2003057146-A2.
 XX
 PD 17-JUL-2003.
 XX
 PF 26-DEC-2002; 2002WO-US041414.
 XX
 PR 26-DEC-2001; 2001US-00035832.
 XX
 PA (SAGR-) SAGRES DISCOVERY.
 XX
 PI Morris DW;
 XX
 DR WPI; 2003-587068/55.
 XX
 PT New recombinant nucleic acid encoding carcinoma associated protein,
 PT useful for preparing compositions for treating carcinomas.
 XX
 PS Claim 1; SEQ ID NO 1280; 245pp; English.
 XX
 CC The invention relates to recombinant carcinoma associated (CA) nucleic
 CC acid sequences from mouse and human (ADA01482-ADA03094), and to
 CC recombinant carcinoma associated proteins (CAP) encoded by them. The
 CC invention also encompasses expression vectors and host cells comprising a
 CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
 CC binds to the protein, and a blotchip comprising CA nucleic acid or
 CC fragments thereof. The sequences of the invention were identified using
 CC oncogenic retroviruses, which insert into the genome of the host organism
 CC at random. Many of these do not carry transduced host oncogenes or
 CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
 CC direct consequence of the effects of proviral integration into host
 CC protooncogenes. The CA nucleic acid sequences can be used to diagnose
 CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
 CC leukaemia) or a propensity to carcinoma by determination of the sequence
 CC of a CA gene, or by determination of CA gene expression in particular
 CC tissues. CA nucleic acids, proteins and antibodies are also useful as
 CC therapeutic agents and in screening and evaluating drug candidates. The
 CC present sequence represents a specifically claimed human CA nucleic acid
 CC sequence of the invention. Note: The complete sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 23107 BP; 5308 A; 6264 C; 6229 G; 5306 T; 0 U; 0 Other;

Query Match 62.7%; Score 18.8; DB 9; Length 23107;
Best Local Similarity 76.7%; Pred. No. 2.1e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CTGGAGCCCTCTCGACTCGAGAGTTCGGCT 30
DB 14873 CTGGAGCCCTCTCGACTCGAGAGTTCGGCT 14902

RESULT 6
ADB72500
ID ADB72500 standard; DNA; 23107 BP.

XX ADB72500;

DT 04-DEC-2003 (first entry)

XX Human Runx3 gene.

XX human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
XX cancer; neoplasm; adenocarcinoma; sarcoma; gene.

OS Homo sapiens.

XX WO2003008583-A2.

XX 30-JAN-2003.

XX 26-DEC-2001; 2001WO-US051291.

XX 02-MAR-2001; 2001US-00798586.

XX 23-OCT-2001; 2001US-00004113.

XX 08-NOV-2001; 2001US-00052482.

XX 30-NOV-2001; 2001US-00997722.

XX 20-DEC-2001; 2001US-00034650.

XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW, Engelhard EK;

XX WPI; 2003-239337/23.

XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,

XX cancer, neoplasm, adenocarcinoma, or sarcomas.

XX Claim 1; SEQ ID NO 328; 2304bp; English.

XX The invention relates to a novel recombinant nucleic acid comprising a

XX nucleotide sequence selected from any of the 660 sequences fully defined

XX in the specification. A polynucleotide of the invention has cytostatic

XX activity, and may have a use in gene therapy, or in a vaccine. The

XX recombinant nucleic acids and polypeptides are useful for treating

XX carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and

XX sarcomas. The present sequence represents a human gene of the invention.

XX Sequence 23107 BP; 5308 A; 6264 C; 6229 G; 5306 T; 0 U; 0 Other;

XX Query Match 62.7%; Score 18.8; DB 10; Length 23107;

XX Best Local Similarity 76.7%; Pred. No. 2.1e+02;

XX Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CTGGAGCCCTCTCGACTCGAGAGTTCGGCT 30

DB 14873 CTGGAGCCCTCTCGACTCGAGAGTTCGGCT 14902

RESULT 7

ADCB5242

ID ADCB5242 standard; DNA; 23107 BP.

XX ADCB5242;

AC ADCB5242;

XX 01-JAN-2004 (first entry)

DT 01-JAN-2004 (first entry)

XX Human Runx3 genomic sequence.

XX Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;

XX secreted; transmembrane; intracellular; ds.

XX Homo sapiens.

XX WO2003045230-A2.

XX 05-JUN-2003.

XX 02-DEC-2002; 2002WO-US038582.

XX 30-NOV-2001; 2001US-00997722.

XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW, Engelhard EK;

XX WPI; 2003-513603/48.

XX New recombinant nucleic acid comprising a nucleotide sequence of any of

XX the carcinoma-associated (CA) genes, useful for screening for drug

XX candidates for diagnosing or treating carcinomas.

XX Claim 1; SEQ ID NO 28; 983bp; English.

XX The invention relates to a recombinant nucleic acid comprising a

XX nucleotide sequence selected from any of the fully defined carcinoma-

XX associated (CA) genes from the 50 tables given in the specification. The

XX CA proteins are secreted, transmembrane or intracellular proteins. The

XX recombinant nucleic acids are useful for screening for drug candidates

XX for diagnosing or treating carcinomas. Sequences given in ADCB5215-

XX ADCB5514 represent CA genes of the invention.

XX Sequence 23107 BP; 5308 A; 6264 C; 6229 G; 5306 T; 0 U; 0 Other;

XX Query Match 62.7%; Score 18.8; DB 10; Length 23107;

XX Best Local Similarity 76.7%; Pred. No. 2.1e+02;

XX Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CTGGAGCCCTCTCGACTCGAGAGTTCGGCT 30

DB 14873 CTGGAGCCCTCTCGACTCGAGAGTTCGGCT 14902

RESULT 8

ADM74357

ID ADM74357 standard; DNA; 23107 BP.

XX ADM74357;

XX 01-JUN-2004 (first entry)

XX Human carcinoma associated (CA) nucleic acid #13.

XX Human; carcinoma associated nucleic acid; CA nucleic acid; gene; ds;

XX carcinoma associated protein; CAP; carcinoma; leukemia; lymphoma;

XX cytostatic.

XX Homo sapiens.

XX US2004072154-A1.

XX 15-APR-2004.

XX 30-NOV-2001; 2001US-00997722.

XX 22-DEC-2000; 2000US-00747377.

XX 02-MAR-2001; 2001US-00798586.

XX (MORR/) MORRIS D W.

Seq	Sequence	3459 BP; 1082 A; 721 C; 678 G; 978 T; 0 U; 0 Other;
Query Match	62.0%;	Score 18.6; DB 2;
Best Local Similarity	84.0%;	Pred. No. 2.2e+02;
Matches	21; Conservative	0; Mismatches 4; Indels 0; Gaps 0
Qy	1	CTGGACCCCTCTCGACTCGAGAGTT 25
Db	1232	CTGGACCCCTTTTCGATGATGATGTT 1208

PI Essioux L;

12

```

Key      location/Qualifiers
primr_bind 7998..7958
/*tag= a
/note= "binds primer 99-27943..rp"
9297..8315
/*tag= B
/note= "binds primer 99-27943-150..misc"
8304..8328
/*tag= C
/bound_moiety= "Probe_99-27943-150"
8316
/*tag= d
/note= "Bifunctional marker A1"
```

```
FT primer_bind complement(8317. .8335)
FT /tag= e
FT /note= "Binds primer 99-27943-150.mis complement"
FT primer_bind complement(8446. .8465)
FT /tag= f
FT /note= "Binds primer 99-27943.pu complement"
FT primer_bind complement(21365. .21385)
FT /tag= g
FT /note= "Binds primer 99-27935.rp"
FT primer_bind complement(21653. .21671)
FT /tag= h
FT /note= "Binds primer 99-27935-193.mis"
FT misc_binding complement(21660. .21684)
FT /tag= i
FT /bound_motety= "Probe_99-27935-193"
FT misc_feature complement(21672)
FT /tag= j
FT /note= "Biallelic marker A2"
FT primer_bind complement(21673. .21691)
FT /tag= k
FT /note= "Binds primer 99-27935-193.mis complement"
FT primer_bind complement(21845. .21864)
FT /tag= l
FT /note= "Binds primer 99-27935.pu complement"
FT primer_bind complement(65463. .65471)
FT /tag= m
FT /note= "Binds primer 8-128.pu"
FT primer_bind complement(65466. .65484)
FT /tag= n
FT /note= "Binds primer 8-128-33.mis"
FT misc_binding complement(65473. .65497)
FT /tag= o
FT /bound_motety= "Probe_8-128-33"
FT misc_feature complement(65485)
FT /tag= p
FT /note= "Biallelic marker A3"
FT primer_bind complement(65486. .65504)
FT /tag= q
FT /note= "Binds primer 8-128-33.mis complement"
FT primer_bind complement(65856. .65874)
FT /tag= r
FT /note= "Binds primer 8-128.rp complement"
FT primer_bind complement(95034. .95053)
FT /tag= s
FT /note= "Binds primer 99-31960.pu"
FT primer_bind complement(95377. .95395)
FT /tag= t
FT /note= "Binds primer 99-31960-363.mis"
FT misc_binding complement(95384. .95408)
FT /tag= u
FT /bound_motety= "Probe_99-31960-363"
FT misc_feature complement(95396)
FT /tag= v
FT /note= "Biallelic marker A4"
FT primer_bind complement(95397. .95415)
FT /tag= w
FT /note= "Binds primer 99-31960-363.mis complement"
FT primer_bind complement(95543. .95563)
FT /tag= x
FT /note= "Binds primer 99-31960.rp complement"
FT primer_bind complement(107022. .107040)
FT /tag= y
FT /note= "Binds primer 99-24656.pu"
FT primer_bind complement(107262. .107280)
FT /tag= z
FT /note= "Binds primer 99-24656-260.mis"
FT misc_binding complement(107269. .107293)
FT /tag= aa
FT /bound_motety= "Probe_99-24656-260"
FT misc_feature complement(107281)
FT /tag= ab
FT /note= "Biallelic marker A5"
FT primer_bind complement(107282. .107300)

FT primer_bind /tag= ac
FT /note= "Binds primer 99-24656-260.mis complement"
FT primer_bind complement(107495. .107513)
FT /tag= ad
FT /note= "Binds primer 99-24656.rp complement"
FT primer_bind complement(160279. .160298)
FT /tag= ae
FT /note= "Binds primer 99-24639.rp"
FT primer_bind complement(160621. .160639)
FT /tag= af
FT /note= "Binds primer 99-24639-163.mis"
FT misc_binding complement(160628. .160652)
FT /tag= ag
FT /bound_motety= "Probe_99-24639-163"
FT misc_feature complement(160640)
FT /tag= ah
FT /note= "Biallelic marker A6"
FT primer_bind complement(160641. .160659)
FT /tag= ai
FT /note= "Binds primer 99-24639-163.mis complement"
FT primer_bind complement(160770. .160787)
FT /tag= aj
FT /note= "Binds primer 99-24634.pu"
FT primer_bind complement(160785. .160802)
FT /tag= ak
FT /note= "Binds primer 99-24639.pu complement"
FT primer_bind complement(160857. .160875)
FT /tag= al
FT /note= "Binds primer 99-24634-108.mis"
FT misc_binding complement(160864. .160888)
FT /tag= am
FT /bound_motety= "Probe_99-24634-108"
FT misc_feature complement(160876)
FT /tag= an
FT /note= "Biallelic marker A7"
FT primer_bind complement(160877. .160895)
FT /tag= ao
FT /note= "Binds primer 99-24634-108.mis complement"
FT primer_bind complement(161240. .161257)
FT /tag= ap
FT /note= "Binds primer 99-24634.rp complement"
FT primer_bind complement(168813. .168830)
FT /tag= aq
FT /note= "Binds primer 99-7652.pu"
FT primer_bind complement(168955. .168973)
FT /tag= ar
FT /note= "Binds primer 99-7652-162.mis"
FT misc_binding complement(168962. .168986)
FT /tag= as
FT /bound_motety= "Probe_99-7652-162"
FT misc_feature complement(168974)
FT /tag= at
FT /note= "Biallelic marker A8"
FT primer_bind complement(168975. .168993)
FT /tag= au
FT /note= "Binds primer 99-7652-162.mis complement"
FT primer_bind complement(169331. .169351)
FT /tag= av
FT /note= "Binds primer 99-7652.rp complement"
FT primer_bind complement(170666. .170686)
FT /tag= aw
FT /note= "Binds primer 99-16100.pu"
FT primer_bind complement(170791. .170809)
FT /tag= ax
FT /note= "Binds primer 99-16100-147.mis"
FT misc_binding complement(170798. .170822)
FT /tag= ay
FT /bound_motety= "Probe_99-16100-147"
FT misc_feature complement(170810)
FT /tag= az
FT /note= "Biallelic marker A9"
FT primer_bind complement(170811. .170829)
FT /tag= ba
```

Query	Query Match	Best Local Similarity	Matches	Conservative	Score	DB	Length	Gaps
Db	1 CTGACCCCTCTCGACTCGAGAGTTCG 28	78.6%	22	0	61.3%	DB 5	319608	0
	8630 CTGAACCCATCTCGATTGGAGTTTACG 8717							
RESULT 13								
AAU52198								
XX	AAU52198 standard; cDNA; 340449 BP.							
AC	AAU52198;							
XX								
DT	22-SEP-2003 (first entry)							
XX								
DE	Human secreted protein genomic DNA coding sequence.							
XX								
KW	Human; gene; ds; secreted protein; chromosome 5; tissue typing;							
KW	secreted protein-related disease; transgenic animal; drug screening;							
KW	pharmacognomic analysts; single nucleotide polymorphism; SNP.							
XX								
OS	Homo sapiens.							
XX								
FT	Key	Location/Qualifiers						
FT	variation	replace(1107, T)						
FT		/*tag= a						
FT	variation	/standard_name= "Single nucleotide polymorphism"						
FT		replace(5479, G)						
FT		/*tag= b						
FT	variation	/standard_name= "Single nucleotide polymorphism"						
FT		replace(5480, G)						
FT		/*tag= c						
FT	variation	/standard_name= "Single nucleotide polymorphism"						
FT		replace(6100, T)						
FT		/*tag= d						
FT	variation	/standard_name= "Single nucleotide polymorphism"						
FT		replace(6122, A)						
FT		/*tag= e						
FT	variation	/standard_name= "Single nucleotide polymorphism"						
FT		replace(6883, G)						
FT		/*tag= f						
FT		/standard_name= "Single nucleotide polymorphism"						

FT	variation	replace(7641, G)
FT		/*tag= g
FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		replace(8493, T)
FT		/*tag= h
FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		replace(9526, A)
FT		/*tag= i
FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		replace(9796, A)
FT		/*tag= j
FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		replace(9868, A)
FT		/*tag= k
FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		replace(10209, C)
FT		/*tag= l
FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		replace(10311, G)
FT		/*tag= m
FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		replace(10831, T)
FT		/*tag= n
FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		replace(12230, C)
FT		/*tag= o
FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		replace(13378, C)
FT		/*tag= p
FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		replace(13423, T)
FT		/*tag= q
FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		replace(13425, G)
FT		/*tag= r
FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		replace(14116, G)
FT		/*tag= s
FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		replace(14508, C)
FT		/*tag= t
FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		replace(14645, A)
FT		/*tag= u
FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		replace(14932, A)
FT		/*tag= v
FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		replace(16212, G)
FT		/*tag= w
FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		replace(16945, T)
FT		/*tag= x
FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		replace(17065, G)
FT		/*tag= y
FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		replace(17442, G)
FT		/*tag= z
FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		replace(17475, C)
FT		/*tag= aa
FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		replace(17844, A)
FT		/*tag= ab
FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		replace(17929, G)
FT		/*tag= ac
FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		replace(18540, C)
FT		/*tag= ad
FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		replace(18592, T)

```
FT /tag= ae
FT /standard_name= "Single nucleotide polymorphism"
FT replace(18616, G)
FT /tag= af
FT /standard_name= "Single nucleotide polymorphism"
FT replace(20491, T)
FT /tag= ag
FT /standard_name= "Single nucleotide polymorphism"
FT replace(22023, G)
FT /tag= ah
FT /standard_name= "Single nucleotide polymorphism"
FT replace(22248, G)
FT /tag= ai
FT /standard_name= "Single nucleotide polymorphism"
FT replace(23118, C)
FT /tag= aj
FT /standard_name= "Single nucleotide polymorphism"
FT replace(23789, A)
FT /tag= ak
FT /standard_name= "Single nucleotide polymorphism"
FT replace(23800, A)
FT /tag= al
FT /standard_name= "Single nucleotide polymorphism"
FT replace(23930, A)
FT /tag= am
FT /standard_name= "Single nucleotide polymorphism"
FT replace(24159, G)
FT /tag= an
FT /standard_name= "Single nucleotide polymorphism"
FT replace(26339, A)
FT /tag= ao
FT /standard_name= "Single nucleotide polymorphism"
FT replace(26508, G)
FT /tag= ap
FT /standard_name= "Single nucleotide polymorphism"
FT replace(26570, A)
FT /tag= aq
FT /standard_name= "Single nucleotide polymorphism"
FT replace(26582, G)
FT /tag= ar
FT /standard_name= "Single nucleotide polymorphism"
FT replace(26693, A)
FT /tag= as
FT /standard_name= "Single nucleotide polymorphism"
FT replace(26884, T)
FT /tag= at
FT /standard_name= "Single nucleotide polymorphism"
FT replace(27320, T)
FT /tag= au
FT /standard_name= "Single nucleotide polymorphism"
FT replace(27339, A)
FT /tag= av
FT /standard_name= "Single nucleotide polymorphism"
FT replace(27542, G)
FT /tag= aw
FT /standard_name= "Single nucleotide polymorphism"
FT replace(28586, A)
FT /tag= ax
FT /standard_name= "Single nucleotide polymorphism"
FT replace(28591, T)
FT /tag= ay
FT /standard_name= "Single nucleotide polymorphism"
FT replace(28599, A)
FT /tag= az
FT /standard_name= "Single nucleotide polymorphism"
FT replace(30857, A)
FT /tag= ba
FT /standard_name= "Single nucleotide polymorphism"
FT replace(31503, T)
FT /tag= bb
FT /standard_name= "Single nucleotide polymorphism"
FT replace(33671, A)
FT /tag= bc
```

```
FT /standard_name= "Single nucleotide polymorphism"
FT replace(35045, C)
FT /tag= bd
FT /standard_name= "Single nucleotide polymorphism"
FT replace(35944, G)
FT /tag= be
FT /standard_name= "Single nucleotide polymorphism"
FT replace(37157, G)
FT /tag= bf
FT /standard_name= "Single nucleotide polymorphism"
FT replace(37603, T)
FT /tag= bg
FT /standard_name= "Single nucleotide polymorphism"
FT replace(39242, G)
FT /tag= bh
FT /standard_name= "Single nucleotide polymorphism"
FT replace(39404, A)
FT /tag= bi
FT /standard_name= "Single nucleotide polymorphism"
FT replace(40395, C)
FT /tag= bj
```

Query Match 60.7%; Score 18.2; DB 8; Length 340449;
Best Local Similarity 87.0%; Pred. No. 4.7e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TGGACCCCTTCGACTCGAGACT 24

DB 25999 TGGACCCCTCCCGACTCGAGACT 26021

RESULT 14

ADB74382 ID ADB74382 standard; DNA; 42325 BP.

AC ADB74382;

DT 04-DEC-2003 (first entry)

XX Mycobacterium leprae DNA #16.

KW Non-naturally occurring peptide; anion pump protein; tuberculosis;

KX hypersensitivity reaction; tuberculostatic; gene; de.

OS Mycobacterium leprae.

XX US6583266-B1.

PD 24-JUN-2003.

PF 16-SEP-1994; 94US-00311731.

XX 19-AUG-1993; 93US-00109181.

PR 22-OCT-1993; 93US-00142558.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Smith DR, Mao J;

PI WPI; 2003-656441/62.

DR New Mycobacterium tuberculosis anion pump peptide useful for as

XX tuberculosis vaccine and diagnosis of tuberculosis infection.

XX Disclousure; SEQ ID NO 131; 26pp; English.

PS The invention relates to a non-naturally occurring peptide of

CC Mycobacterium tuberculosis comprising an amino acid sequence

CC corresponding to an anion pump protein. The invention also relates to a

CC non-naturally occurring nucleic acid corresponding to a DNA sequence of

CC Mycobacterium tuberculosis or Mycobacterium leprae. The new peptide is

CC useful as a vaccine against Mycobacterium tuberculosis or Mycobacterium

CC leprae or for screening for new tuberculosis drugs. Purified proteins

CC derived from the sequences of the invention may elicit a specific immune
CC response. The peptide may also be used to detect hypersensitivity
CC reactions of individuals exposed to Mycobacterium tuberculosis or
CC Mycobacterium leprae. The proteins and peptides may be affixed to solid
CC supports to detect antibodies typical of hypersensitivity reactions, from
CC a patient's sera. This sequence represents Mycobacterium leprae DNA of
CC the invention. Note: The sequence data for this patent did not form part
CC of the printed specification but was obtained in electronic format
CC directly from USPTO at seqdata.uspto.gov/sequence.html.
XX

SQ Sequence 42325 BP; 9673 A; 13128 C; 11330 G; 8194 T; 0 U; 0 Other;

Query Match 60.0%; Score 18; DB 10; Length 42325;
Best Local Similarity 80.8%; Pred. No. 5e+02;

Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 ACCCTCTGACTCGAGATTCCGCT 30
DB 688 ACCCACTGACTCCAGTACGCT 713

RESULT 15

ADM27081_10/c

Continuation (11 of 17) of ADM27081 from base 1000001 (Hyperthermophile Methanopyrus kar
WP Sequence split into 17 fragments LOCUS ADM27081 Accession Adm27081

WP	Fragment Name	Begin	End
WP	ADM27081_00	1	110000
WP	ADM27081_01	100001	210000
WP	ADM27081_02	200001	310000
WP	ADM27081_03	300001	410000
WP	ADM27081_04	400001	510000
WP	ADM27081_05	500001	610000
WP	ADM27081_06	600001	710000
WP	ADM27081_07	700001	810000
WP	ADM27081_08	800001	910000
WP	ADM27081_09	900001	1010000
WP	ADM27081_10	1000001	1110000
WP	ADM27081_11	1100001	1210000
WP	ADM27081_12	1200001	1310000
WP	ADM27081_13	1300001	1410000
WP	ADM27081_14	1400001	1510000
WP	ADM27081_15	1500001	1610000
WP	ADM27081_16	1600001	1694968

Query Match 60.0%; Score 18; DB 11; Length 110000;
Best Local Similarity 80.8%; Pred. No. 5.4e+02;

Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CTGAGCCCTCTCGACTCGAGATTTC 26
DB 69684 CTTGCCGCTCTGAGCTCGAGATTTC 69659

Search completed: February 28, 2005, 23:51:44
Job time : 268.818 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 28, 2005, 23:32:36 ; Search time 875.065 Seconds
(without alignments)
1661.198 Million cell updates/sec

Title: US-10-086-062-4

Perfect score: 30
Sequence: 1 ctgagccctctcgactcgagagttccgct 30

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_hgt: *
3: gb_in: *
4: gb_ov: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_srs: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	100.0	30	6 AX358109	AX358109 Sequence
2	20.4	68.0	595	9 HUMQ76F01	AP085912 Homo sapi
3	20.4	68.0	98348	9 AL136103	AL136103 Human DNA
4	20.4	68.0	178965	9 AC012048	AC012048 Homo sapi
5	20.2	67.3	23211	3 AC099767	AC099767 Caenorhab
6	20	66.7	1503	1 AB038027	AB038027 Vibriol sp
7	20	66.7	156165	2 AC021453	AC021453 Homo sapi
8	20	66.7	157980	2 AC100852	AC100852 Homo sapi
9	19.6	65.3	25785	2 AC123011	AC123011 Rattus no
10	19.4	64.7	11406	4 AF107201	AF107201 Equus cab
11	19.4	64.7	70387	3 AC087075	AC087075 Caenorhab
12	19.4	64.7	110000	8 AC104791	AC104791 Homo sapi
13	19.4	64.7	159969	9 AC104791	Continuation (7 of
14	19.4	64.7	189269	2 AP006440	AP006440 Homo sapi
15	19.4	64.7	196472	2 AC011862	AC011862 Homo sapi
16	19.2	64.0	1500	1 AF118021	AF118021 Vibriol ca
17	19.2	64.0	1500	6 AR452360	AR452360 Sequence
18	19.2	64.0	1500	6 AX201179	AX201179 Sequence
19	63.3	93695	8	ATP2009	AL021749 Arabidops

c 20	19	63.3	156806	2 AC021353	AC021353 Homo sapi
c 21	19	63.3	177652	10 AC140251	AC140251 Mus muscu
c 22	19	63.3	199749	8 ATCHRIV68	AL161572 Arabidops
c 23	19	63.3	203120	10 AC124991	AC124991 Mus muscu
c 24	19	63.3	205475	2 AC130033	AC130033 Rattus no
c 25	19	63.3	213906	2 AC118634	AC118634 Mus muscu
c 26	19	63.3	223675	9 AC025031	AC025031 Homo sapi
c 27	19	63.3	232320	10 AC127314	AC127314 Mus muscu
c 28	19	63.3	244974	2 AC111249	AC111249 Rattus no
c 29	18.8	62.7	2481	6 AX827828	AX827828 Sequence
c 30	18.8	62.7	2481	10 R4TR11BA2	W23366 Rattus norv
c 31	18.8	62.7	5550	1 AB016787	AB016787 Pseudomon
c 32	18.8	62.7	23107	6 AX695653	AX695653 Sequence
c 33	18.8	62.7	35758	2 AC127539	AC127539 Homo sapi
c 34	18.8	62.7	67145	2 AC124281	AC124281 Homo sapi
c 35	18.8	62.7	69833	2 AC010777	AC010777 Homo sapi
c 36	18.8	62.7	77691	9 AC120118	AC120118 Homo sapi
c 37	18.8	62.7	95241	9 HS39819	AL023036 Human DNA
c 38	18.8	62.7	100521	9 AC119397	AC119397 Homo sapi
c 39	18.8	62.7	101721	9 AC067950	AC067950 Homo sapi
c 40	18.8	62.7	110000	8 CR882131_15	Continuation (16 o
c 41	18.8	62.7	113196	9 HS3697K14	AL121829 Human DNA
c 42	18.8	62.7	141605	2 AC013732	AC013732 Homo sapi
c 43	18.8	62.7	146606	9 AC068213	AC068213 Homo sapi
c 44	18.8	62.7	148332	9 AC079586	AC079586 Homo sapi
c 45	18.8	62.7	148611	8 AC091680	AC091680 Oryza sat

ALIGNMENTS

RESULT 1
AX358109
LOCUS AX358109 Sequence 4 from Patent WO0194394. 30 bp DNA linear PAT 13-FEB-2002

ACCESSION AX358109
VERSION AX358109.1 GI:18674856

KEYWORDS
SOURCE Agrobacterium tumefaciens (Rhizobium radiobacter)
ORGANISM Agrobacterium tumefaciens
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.

REFERENCE
1 Jilka, J.M., Hood, B.E. and Howard, J.A.

AUTHORS
TITLE Novel plant promoter sequences and methods of use for same
JOURNAL Patent: WO 0194394-A 4 13-DEC-2001;
Prodigene, Inc. (US)

FEATURES
source
1..30
/organism="Agrobacterium tumefaciens"
/mol_type="unassigned DNA"
/db_xref="taxon:358"

ORIGIN

Query Match 100.0%; Score 30; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGAGACCCCTCTCGACTCGAGAGTTCCGCT 30
Db 1 CTGAGACCCCTCTCGACTCGAGAGTTCCGCT 30

RESULT 2
HUMQ76F01/c HUMQ76F01 595 bp mRNA linear PRI 29-AUG-1998

LOCUS HUMQ76F01 Homo sapiens full length insert cDNA clone YQ76F01.
DEFINITION AP085912

ACCESSION AP085912
VERSION AP085912.1 GI:3463252
KEYWORDS FLI_CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE
1 (bases 1 to 595)	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
Woessner,B., Tan,F.,	Marra,M., Kucaba,T., Vandell,M., Martin,J.,			
Math,G., Bowles,L.,	Wylie,T., Bowers,Y., Stepec,M., Theising,B.,			
Getseil,S., Allen,M.,	Underwood,K., Chappell,J., Person,B.,			
Gibbons,M., Harvey,N.,	Pape,D., Chamberlain,A., Morales,R.,			
Schurk,R., Ritzer,E.,	Kohn,S., Swaller,T., Behmer,K., Hillier,L.,			
Wilson,R. and Waterston,R.				
Full Clone Sequencing of the Longest Available Member from Each				
Unpublished				
2 (bases 1 to 595)				
Waterston,R.				
Direct Submission				
Submitted (24-AUG-1998)	Department of Genetics, Washington			
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA				
SUBMITTED BY:				
Genome Sequencing Center				
Department of Genetics				
Washington University				
St. Louis MO 63108, USA				
http://genome.wustl.edu/gsc				
mailto:est@watson.wustl.edu				

NOTICE: This sequence represents the full insert of this cDNA. No attempt has been made to verify whether this corresponds to the full-length of the original mRNA from which it was derived. We have tried to obtain double-stranded, or double chemistry sequence across the entire clone, but potentially, there are areas in the sequence where this level of coverage was not achieved. Nevertheless, we are confident of the accuracy of this sequence as all regions of low quality, as defined by PHRAP (P. Green, in preparation), were visually inspected and edited accordingly. The consensus quality values for this sequence have been submitted separately.

The location of this clone is unknown.

```

features
  source
    location/Qualifiers
      1..595
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="IMAGE:201721"
        /clone_1kb="Scores_fetal_liver_spleen_1MPLS
repeat_region
  9..338

```

ORIGIN

Query Match	68.0%	Score 20.4	DB 9	Length 595
Best Local Similarity	80.0%	Pred. No. 2.9e+02		
Matches 24; Conservative	0	Mismatches 6	Indels 0	Gaps 0

Dy 1 CTGACCCCTTCGACTCGAGATTCCGCT 30
Dd 566 CTGACCCCTTCGCTTGAGACTTCCCTCT 537

RESULT 3	AL136103	98348 bp	DNA	linear	PRI 04-JAN-2007
LOCUS	AL136103				
DEFINITION	Human DNA sequence from clone RPl-250B11 on chromosome 10 Contains STSs and GSSs, complete sequence.				
ACCESSION	AL136103				

SOURCE ORGANISM	REFERENCE
Homo sapiens (human)	
Homo sapiens	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	1 (bases 1 to 98348)

REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 98348)
Phillimore, B.
Direct Submission
Submitted (04-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire

COMMENT

CB10 ISA. E-mail enquiries: hunquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Jul 28, 2000 this sequence version replaced gi:9501151.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences without a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wb, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep

This sequence is the entire insert of clone RPL-250B11. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unure' feature key. This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr10>

RPL-250B11 is from the library RPCT-1 constructed by the group of Pletcher de Jong. For further details see <http://www.choir1.org/bacpac/home.htm>

VECTOR: pCYPAC2.

FEATURES

Source

misc_feature

repeat_region

repeat_region

repeat_region

100-44388-1

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

—

3. 2004. 10. 10.

misc feature

repeat region

repeat_region

repeat_region

repeat_region

100-443887-1

1
2
3
4

```

USA, UK, E-nil enquiries: humquerry@sanger.ac.uk Clone
URL: clonerequest@sanger.ac.uk
28, 2000 this sequence version replaced g1:9501151.
sequence assembly data is compared from overlapping clones
differences are found these are annotated as variations
with a note of the overlapping clone name. Note that the
on annotation may not be found in the sequence submission
according to the overlapping clone, as we submit sequences with
small overlap as described above.
allowing abbreviations are used to associate primary accessions
as given in the feature table with their source databases:
WORMPEP, Sw., SWISSPROT, Tr., TREMBL, Wp., WORMPEP, Information
WORMPEP database can be found at
www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
fire insert of clone RP1-250B11 This sequence has been
and according to sequence map criteria as follows. An attempt
to resolve all sequencing problems, such as compressions
delets, but not necessarily within known annotated repeat
elements. Where the sequence is ambiguous, there is an
tion using the 'unsure' feature key. This sequence was
formed from part of bacterial clone contigs of human chromosom
structured by the Sanger Centre chromosome 10 Mapping Group.
Information can be found at
www.sanger.ac.uk/Help/Chr10
DBI is from the library RPc1-1 constructed by the group of
de Jong. For further details see
www.chori.org/bacpac/home.htm
PCYPAC2.
Location/Qualifiers
1. 98348
organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP1-250B11"
/clone_1lb="RPc1-1"
1. 334
/note="Single clone region. Reads derived from clone PCR
Assembly consistent with restriction digest."
2458. 2765
/note="AluXs repeat: matches 1. 308 of consensus"
3087. 3267
/note="MER3b repeat: matches 355. 546 of consensus"
3269. 3506
/note="MER3 repeat: matches 13. 380 of consensus"
6106. 6337
/note="L2 repeat: matches 2435. 2705 of consensus"
6344. 6662
/note="AluSg repeat: matches 1. 298 of consensus"
complement(8333. 8799)
/note="match: GSS: Em:AQ403443"
complement(8375. 8758)
/note="match: GSS: Em:AQ264393"
8802. 9234
/note="match: GSS: Em:AQ150662"
9268. 9356
/note="match: STS: Em:G28019"
11809. 12092
/note="AluXs repeat: matches 1. 284 of consensus"
12205. 12722
/note="LIME repeat: matches 5285. 5811 of consensus"
14113. 14171
/note="L2 repeat: matches 2692. 2750 of consensus"
14548. 14846
/note="AluY repeat: matches 1. 299 of consensus"
15883. 16169
/note="AluXs repeat: matches 1. 287 of consensus"
17146. 18557
/note="MER52A repeat: matches 1. 1755 of consensus"
18559. 18713
/note="MIR repeat: matches 6. 160 of consensus"
18935. 19381
/note="MER63 repeat: matches 590. 1061 of consensus"

```



```
repeat_region 19384..19519
/note="FLAM C repeat: matches 3. .137 of consensus"
repeat_region 19713..19836
/note="MER63 repeat: matches 463. .577 of consensus"
repeat_region 19826..20117
/note="MER63 repeat: matches 2. .307 of consensus"
repeat_region 20325..20391
/note="MIR repeat: matches 88. .156 of consensus"
repeat_region 20440..20701
/note="MIR repeat: matches 2. .248 of consensus"
misc_feature 22321..22631
/note="match: STR: Em:HSJ10C4"
misc_feature 25725..26188
/note="match: GSS: Em:AQ211002"
misc_feature 25738..25978
/note="match: GSS: Em:AQ823705"
misc_feature 26548..27067
/note="match: GSS: Em:AQ798605"
misc_feature 26564..27250
/note="match: GSS: Em:AQ055423"
misc_feature 27549..27913
/note="match: GSS: Em:B87793"
repeat_region 28724..28952
/note="L2 repeat: matches 2167. .2416 of consensus"
misc_feature complement(29565..30037)
/note="match: GSS: Em:AQ155973"
misc_feature 30073..30670
/note="match: GSS: Em:AQ195052"
misc_feature complement(31752..32121)
/note="match: GSS: Em:AQ099305"
misc_feature complement(32344..32928)
/note="match: GSS: Em:AQ485233"
misc_feature 32930..33368
/note="match: GSS: Em:AQ424293"
misc_feature 34554..34583
/note="L5 copies 2 mer ca 100% conserved"
repeat_region 35059..35135
/note="L2 repeat: matches 2417. .2500 of consensus"
misc_feature complement(35681..36191)
/note="match: GSS: Em:B55923"
misc_feature complement(35777..36102)
/note="match: GSS: Em:AQ223744"
misc_feature complement(35861..36186)
/note="match: GSS: Em:AQ803876"
repeat_region 36692..36800
/note="MIR repeat: matches 13. .123 of consensus"
repeat_region 40461..40765
/note="AluX repeat: matches 1. .293 of consensus"
repeat_region 41312..41502
/note="MIR repeat: matches 8. .192 of consensus"
repeat_region 41785..41840
/note="28 copies 2 mer tg 85% conserved"
repeat_region 43022..43334
/note="AluX repeat: matches 3. .312 of consensus"
misc_feature complement(44477..45066)
/note="match: GSS: Em:AQ540223"
misc_feature complement(44671..45063)
/note="match: GSS: Em:AQ337658"
repeat_region 45077..45157
/note="MER57-internal repeat: matches 7151. .7230 of consensus"
repeat_region 45158..45304
/note="L1P repeat: matches 5050. .5197 of consensus"
repeat_region 45284..45621
/note="L1P4 repeat: matches 5476. .5812 of consensus"
misc_feature complement(45941..46452)
/note="match: GSS: Em:AQ832143"
repeat_region 48564..48709
/note="L2 repeat: matches 2596. .2749 of consensus"
repeat_region 48720..49082
/note="TH1B repeat: matches 1. .364 of consensus"
repeat_region 49176..49629
/note="MLTID repeat: matches 1. .502 of consensus"
```

```
repeat_region 49784..49912
/note="FLAM A repeat: matches 8. .132 of consensus"
repeat_region 49914..50164
/note="MIR repeat: matches 2. .261 of consensus"
repeat_region 52850..52918
/note="L2 repeat: matches 2635. .2706 of consensus"
misc_feature 53092..53613
/note="match: GSS: Em:AQ800606"
misc_feature 53110..53608
/note="match: GSS: Em:AQ172250"
misc_feature 53151..53604
/note="match: GSS: Em:AQ170181"
misc_feature 53165..53586
/note="match: GSS: Em:AQ545768"
misc_feature 53179..53355
/note="match: GSS: Em:AQ898990"
misc_feature 53378..53834
/note="match: GSS: Em:AQ898990"
repeat_region 53776..53958
/note="MIR repeat: matches 41. .250 of consensus"
misc_feature complement(53900..54343)
/note="match: GSS: Em:AQ400632"
repeat_region 53959..54203
/note="AluX repeat: matches 65. .308 of consensus"
repeat_region 54204..54216
/note="MIR repeat: matches 250. .260 of consensus"
misc_feature 54361..54767
/note="match: GSS: Em:AQ729256"
repeat_region 54372..54565
/note="MIR repeat: matches 13. .261 of consensus"
misc_feature 54440..54730
/note="match: GSS: Em:AQ084956"
misc_feature complement(55008..55535)
/note="match: GSS: Em:AQ556262"
misc_feature 55610..55935
/note="match: GSS: Em:AQ594863"
repeat_region 56871..57293
/note="L1MD repeat: matches 883. .1319 of consensus"
repeat_region 57297..57602
/note="AluX repeat: matches 1. .305 of consensus"
```

```
Query Match 68.0%; Score 20.4; DB 9; Length 98348;
Best Local Similarity 80.0%; Pred.No.2e+02;
Matches 24; Mismatches 6; Indels 0; Gaps 0;

QY 1 CTGACCCCTCTCGACTCGAGGTTCCGCT 30
Db 62965 CTGAGCCCTCTCGGCTGAGACTTCCTCT 62994

RESULT 4
LOCUS AC012048 178965 bp DNA linear PRI 20-APR-2002
DEFINITION Homo sapiens chromosome 10 clone RP11-43N22, complete sequence.
ACCESSION AC012048
VERSION AC012048.11 GI:19744964
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 178965)
REFERENCE 1 Smith,D.R.
AUTHORS Smith,D.R.
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome
Sequence Data
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 178965)
AUTHORS Smith,D.R.
TITLE Direct Substitution
JOURNAL Submitted (19-OCT-1999) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
3 (bases 1 to 178965)
REFERENCE Smith,D.R.
```

TITLE Direct Submission
JOURNAL Submitted (16-AUG-2001) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
REFERENCE 4 (bases 1 to 178965)
AUTHORS Smith, D.R.
TITLE Direct Submission
JOURNAL Submitted (27-MAR-2002) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
REFERENCE 5 (bases 1 to 178965)
AUTHORS Smith, D.R.
TITLE Direct Submission
JOURNAL Submitted (28-MAR-2002) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
REFERENCE 6 (bases 1 to 178965)
AUTHORS Smith, D.R.
TITLE Direct Submission
JOURNAL Submitted (20-APR-2002) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
COMMENT On Mar 27, 2002 this sequence version replaced gi:15193325.
FEATURES
source
Location/Qualifiers
1..178965
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-43N22"
/clone_11b="RPCT-11"
ORIGIN
Query Match 68.0%; Score 20.4; DB 9; Length 178965;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 CTGAGACCCCTCTCGACTCGAGAGTTCCGCT 30
Db 32913 CTGAGACCCCTCTCGACTCGAGAGTTCCGCT 32942
RESULT 5
AC099767/c 23211 bp DNA linear INV 20-NOV-2001
LOCUS
DEFINITION Caenorhabditis briggsae cosmid G24F01, complete sequence.
ACCESSION AC099767
VERSION AC099767.1 GI:17017639
KEYWORDS HTG.
SOURCE Caenorhabditis briggsae
ORGANISM Caenorhabditis briggsae
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditioidea; Rhabditidae; Pelodierinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 23211)
AUTHORS Washington University Genome Sequencing Center.
TITLE Washington University Genome Sequencing Project.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 23211)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 23211)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (20-NOV-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 444 Forest Park Avenue, St.
Louis, MO 63110, USA
COMMENT Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63110, USA
email: tw@nematoe.wustl.edu
NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate chemistry
or covered by high quality data (i.e., phred quality >= 30); an
attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by sequence from
more than one m3 subclone.
Location/Qualifiers
1..23211
/organism="Caenorhabditis briggsae"
/mol_type="genomic DNA"
/strain="Gujarat G16"
/db_xref="taxon:6238"
ORIGIN
Query Match 67.3%; Score 20.2; DB 3; Length 23211;
Best Local Similarity 88.0%; Pred. No. 2.7e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 4 GACCCCTCTCGACTCGAGAGTTCCG 28
Db 14462 GATTCCTCTCACTCGAGAGTTCCG 14438
RESULT 6
AB038027 1503 bp DNA linear BCT 18-FEB-2000
LOCUS
DEFINITION Vibrio sp. SG128 gene for 16S rRNA, partial sequence.
ACCESSION AB038027
VERSION AB038027.1 GI:7007473
KEYWORDS 16S ribosomal RNA.
SOURCE Vibrio sp. SG128
ORGANISM Vibrio sp. SG128
Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
Vibrionaceae; Vibrrio.
REFERENCE 1 (bases 1 to 1503)
AUTHORS Urakawa, H.
TITLE 16S rRNA gene of marine bacterium
JOURNAL Published Only in Database (2000)
REFERENCE 2 (bases 1 to 1503)
AUTHORS Urakawa, H.
TITLE Direct Submission
JOURNAL Submitted (05-FEB-2000) Hidetoshi Urakawa, Northwestern University,
Department of Civil Engineering; Technological Institute 2145
Sheridan Road, Evanston, Illinois 60208-3109, USA
(E-mail: h-urakawa@nwu.edu, Tel: +1-847-467-5710,
Fax: +1-847-491-4011)
FEATURES
source
Location/Qualifiers
1..1503
/organism="Vibrio sp. SG128"
/mol_type="genomic DNA"
/strain="SG128"
/db_xref="taxon:115126"
<1..>1503
/product="16S ribosomal RNA"
ORIGIN
Query Match 66.7%; Score 20; DB 1; Length 1503;
Best Local Similarity 82.1%; Pred. No. 4.1e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 CTGAGACCCCTCTCGACTCGAGAGTTCCG 28
Db 231 CTGGGCCCATCCGACGCGAGAGGTCCG 204
RESULT 7
AC021453/c 156165 bp DNA linear HTG 01-APR-2000
LOCUS
DEFINITION Homo sapiens clone RP11-125C16, WORKING DRAFT SEQUENCE, 14
unordered pieces.
ACCESSION AC021453
VERSION AC021453.3 GI:7382318

KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens (human)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 156165)
 TITLE Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 JOURNAL Homo sapiens, clone RP11-125C16
 REFERENCE 2 (bases 1 to 156165)
 TITLE Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A., Choepl, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Deaellano, K., Dewar, K., Domino, M., Doyle, M., Feneator, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galegan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Lehoczk, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marguis, N., McEwan, P., McGurk, A., McKernan, K., McPheters, R., Meldrim, J., Menus, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Olivar, T. M., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
 DIRECT SUBMISSION
 Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 1, 2000 this sequence version replaced g1:6721267.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L5169
 Center clone name: 125_C16
 ----- Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 145142 bases at least Q40
 Consensus quality: 151009 bases at least Q30
 Consensus quality: 153345 bases at least Q20
 Insert size: 16000; agarose-fp
 Insert size: 154865; sum-of-contents
 Quality coverage: 4.1 in Q20 bases; agarose-fp
 Quality coverage: 4.2 in Q20 bases; sum-of-contents

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 14 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 1836: contig of 1836 bp in length
 * 1837 1936: gap of 100 bp
 * 1937 5796: contig of 3860 bp in length
 * 5797 5896: gap of 100 bp
 * 5897 9797: contig of 3901 bp in length
 * 9798 9897: gap of 100 bp
 * 9898 13919: contig of 4022 bp in length
 * 13920 14019: gap of 100 bp
 * 14020 17400: contig of 3381 bp in length
 * 17401 17500: gap of 100 bp
 * 17501 21253: contig of 3753 bp in length

FEATURES
 source
 1..156165
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="RP11-125C16"
 /clone_1ib="RP11-125C16 Human Male BAC"
 1..1836
 /note="assembly_fragment"
 1937..5796
 /note="assembly_fragment"
 5897..9797
 /note="assembly_fragment"
 9898..13919
 /note="assembly_fragment"
 14020..17400
 /note="assembly_fragment"
 17501..21253
 /note="assembly_fragment"
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:right"
 21354..31096
 /note="assembly_fragment"
 31197..41465
 /note="assembly_fragment"
 41566..50901
 /note="assembly_fragment"
 51002..62041
 /note="assembly_fragment"
 62142..75247
 /note="assembly_fragment"
 75348..89874
 /note="assembly_fragment"
 /note="assembly_fragment"
 clone_end:T7
 vector_side:right"
 89975..116362
 /note="assembly_fragment"
 116463..156165
 /note="assembly_fragment"
 ORIGIN
 Query Match 66.7%; Score 20; DB 2; Length 156165;
 Best Local Similarity 82.1%; Pred. No. 2.9e+02;
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 2 TGGACCCCTCGACTCGAGAGTTCGGC 29
 Db 11285 TGGACCTCTCTATTCAAGAGTCTGC 11258
 RESULT 8
 AC100852/c 157980 bp DNA linear PRI 29-AUG-2002
 LOCUS AC100852/c
 DEFINITION Homo sapiens chromosome 17, clone RP11-125C16, complete sequence.
 ACCESSION AC100852
 VERSION AC100852.2 GI:22539166
 KEYWORDS HTG.

```

SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 157980)
AUTHORS     Birren,B., Nussbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome 17, clone RP11-125C16
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 157980)
AUTHORS     Birren,B., Linton,L., Nussbaum,C., Lander,E., All,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouknight,B.,
            Brown,A., Camarata,J., Campiolo,M., Collins,S., Collamore,A., Cook,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
            Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S.,
            Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gaidyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hago,B., Heaford,A., Horton,L., Hulme,W., Illiev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kelle,C., Lacroque,K.,
            Lamazeres,R., Landers,T., Lehotzky,J., Levine,R., Liu,G.,
            Maclean,C., Macdonald,P., Major,J., Margolis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPherson,R., Meldrum,J.,
            Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Noru,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
            Oliver,J., Peterson,K., Phukhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,W., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Teefaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zemke,L., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
REFERENCE   Research, 320 Charles Street, Cambridge, MA 02141, USA
AUTHORS     3 (bases 1 to 157980)
            Birren,B., Nussbaum,C., Lander,E., All,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Bouknight,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A.,
            Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Fero,S., Ferreira,P., Fitzgerld,M., Gage,D., Galagan,J.,
            Gaidyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hago,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kelle,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrum,J., Menus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
            Phukhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Teefaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zemke,L., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (29-AUG-2002) Whitehead Institute/MIT Center for Genome
COMMENT     Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Aug 29, 2002 this sequence version replaced gi:17046222.
            All repeats were identified using RepeatMasker:
            Smit, A.P.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -----
            Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            -----
            Project Information
            Center project name: L21587
            Center clone name: 125_C_16
            -----
FEATURES
source
location/Qualifiers
1..157980
/organism="Homo sapiens"
/mol_type="genomic DNA"

```

```

/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-125C16"
/clone_1lb="RP11-125C16"
62..173
/rpt_family="MIR"
repeat_region
complement(636..832)
/rpt_family="MIR"
repeat_region
833..888
/rpt_family="MER3"
complement(891..1207)
/rpt_family="AluDb"
repeat_region
1208..1283
/rpt_family="MER3"
1232..1236
/nc="< 30 qual SINGL region"
complement(1303..2370)
/rpt_family="L1PA4"
1330..1354
/nc="<30 qual SINGL region"
2371..3213
/rpt_family="L1PA4"
complement(3266..3408)
/rpt_family="Charlie7"
3506..3538
/rpt_family="L2"
3539..3653
/rpt_family="MER47A"
complement(3654..3926)
/rpt_family="AluSc"
3927..4141
/rpt_family="MER47A"
4142..4438
/rpt_family="AluY"
4439..4473
/rpt_family="MER47A"
4474..4561
/rpt_family="L2"
complement(4601..4700)
/rpt_family="L2"
complement(4708..4845)
/rpt_family="MER5A"
complement(4944..5104)
/rpt_family="MIR"
complement(5112..5532)
/rpt_family="L2"
complement(5546..5822)
/rpt_family="AluDb"
complement(5823..5999)
/rpt_family="L2"
6000..6216
/rpt_family="AluDb"
complement(6217..6292)
/rpt_family="L2"
6447..6752
/rpt_family="AluDb"
7412..7472
/rpt_family="A-rich"
7789..7907
/rpt_family="CGG)n"
8642..8683
/rpt_family="T-rich"
8954..8978
/rpt_family="CCA)n"
9084..9123
/rpt_family="CAG)n"
10990..11017
/rpt_family="AT-rich"
11225..11326
/rpt_family="L2"
complement(11372..11504)
/rpt_family="MIR1B"

```

```

repeat_region 13034..13209
                /rpt_family="MIR"
repeat_region complement(15480..15590)
                /rpt_family="L2"
repeat_region 15703..15968
                /rpt_family="AluSg"
repeat_region complement(16005..16318)
                /rpt_family="L2"
repeat_region 16413..16555
                /rpt_family="MER106B"
repeat_region 16825..17099
                /rpt_family="Aluub"
repeat_region 17241..17261
                /rpt_family="AT_rich"
repeat_region 17280..17310
                /rpt_family="(TTG)n"
repeat_region complement(17366..17658)
                /rpt_family="AluSc"
repeat_region 17707..17808
                /rpt_family="MIR"
repeat_region 18250..18277
                /rpt_family="AT_rich"
repeat_region complement(18576..18864)
                /rpt_family="MER3"
repeat_region complement(18926..19077)
                /rpt_family="AluJo"
repeat_region 20786..20952
                /rpt_family="AluJo"
repeat_region 20962..21087
                /rpt_family="MIR"
repeat_region complement(21088..21390)
                /rpt_family="AluSc"
repeat_region complement(21419..21646)
                /rpt_family="AluJo"
repeat_region 21647..21715
                /rpt_family="MIR"
repeat_region complement(21969..22223)
                /rpt_family="L1ME4A"
repeat_region complement(22288..22411)
                /rpt_family="WIR3"
repeat_region complement(22950..23356)
                /rpt_family="AluSg"
repeat_region 24607..24789
                /rpt_family="FAM"
repeat_region 24846..25144
                /rpt_family="Aluub"
repeat_region 25229..25254
                /rpt_family="AT_rich"
repeat_region 25700..26104

```

```

Query Match 66.7%; Score 20; DB 9; Length 157980;
Best Local Similarity 82.1%; Pred. No. 2.9e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Db 130676 TGGACCTCTCTCTATTCAGAGATTCTGC 130649
2 TGGACCCCTCTGACCTCGAGAGTTCCGC 29
|||||

```

```

RESULT 9
AC123011 257595 bp DNA linear HTG 12-OCT-2002
LOCUS Rattus norvegicus clone CH230-92124, WORKING DRAFT SEQUENCE, 4
DEFINITION
AC123011
AC123011.3 GI:23665280
VERSION HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 257595)

```

AUTHORS

Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angiano, D., Anyalebechi, V., Ayogbi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benamed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Butrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cessari, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Day-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Becotto, M., Eugene, C., Evans, C., A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabsis, A., Gantla, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, B., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S., L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Hulik, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C., L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshew, L., Louissege, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundaes, M., Murphy, M., Naif, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackelme, O., Okmonu, G., Olarunbasogun, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C., Ploper, F., Poindecker, A., Popovic, D., Primm, E., Pu, L., L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M., A., Reish, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, J., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C. D., Smjs, D., Sneed, A., Sodergren, E., Song, X., Z., Sorrelle, R., Sosa, D., Stealmle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usman, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

TITLE

Journal
 Unpublished
 2 (bases 1 to 257595)

AUTHORS

Worley, K. C.
 Direct Submission
 Submitted (26-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 257595)

REFERENCE

Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (12-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Oct 10, 2002 this sequence version replaced gi:21909149.

COMMENT

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.bsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole

genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine
Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GYMD
Center clone name: CH230-92124

----- Summary Statistics

Assembly program: Phrap; version 0.990329
Consensus quality: 236342 bases at least Q40
Consensus quality: 239870 bases at least Q30
Consensus quality: 242005 bases at least Q20
Estimated insert size: 243903; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 250677: contig of 250677 bp in length
* 250678 250777: gap of unknown length
* 250778 252307: contig of 1530 bp in length
* 252308 252407: gap of unknown length
* 252408 253577: contig of 1170 bp in length
* 253578 253678: gap of unknown length
* 253678 257595: contig of 3918 bp in length.

FEATURES

source

1..257595

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-92124"

/complement(246343..247143)

/note="clone boundary"

clone end: Sp6

site: EcoRI

end_sequence: BH295371"

complement(246398..247079)

/note="clone boundary"

clone end: Sp6

site: EcoRI

end_sequence: BH295371"

complement(246398..247079)

/note="clone boundary"

clone end: Sp6

site: EcoRI

end_sequence: BH295371"

complement(246398..247079)

/note="clone boundary"

clone end: Sp6

site: EcoRI

end_sequence: BH295371"

complement(246398..247079)

/note="clone boundary"

clone end: Sp6

site: EcoRI

end_sequence: BH295371"

complement(246398..247079)

/note="clone boundary"

REFERENCE
1 (bases 1 to 11406)
AUTHORS
Maeel,A.M., Brandon,R.B. and Bell,T.K.
TITLE
Nucleotide Sequence of the Equine Beta-lactoglobulin Gene
JOURNAL
Unpublished
AUTHORS
Maeel,A.M., Brandon,R.B. and Bell,T.K.
TITLE
Direct Submission
JOURNAL
Submitted (17-NOV-1998) Australian Equine Blood Typing Research
Laboratory, University of Queensland, St. Lucia, Brisbane,
Queensland 4072, Australia

REFERENCE
1 (bases 1 to 11406)
AUTHORS
Maeel,A.M., Brandon,R.B. and Bell,T.K.
TITLE
Nucleotide Sequence of the Equine Beta-lactoglobulin Gene
JOURNAL
Unpublished
AUTHORS
Maeel,A.M., Brandon,R.B. and Bell,T.K.
TITLE
Direct Submission
JOURNAL
Submitted (17-NOV-1998) Australian Equine Blood Typing Research
Laboratory, University of Queensland, St. Lucia, Brisbane,
Queensland 4072, Australia

FEATURES

source

1..11406

/organism="Equus caballus"

/mol_type="genomic DNA"

/db_xref="taxon:9796"

<4472..>8679

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

REFERENCE
1 (bases 1 to 11406)
AUTHORS
Maeel,A.M., Brandon,R.B. and Bell,T.K.
TITLE
Nucleotide Sequence of the Equine Beta-lactoglobulin Gene
JOURNAL
Unpublished
AUTHORS
Maeel,A.M., Brandon,R.B. and Bell,T.K.
TITLE
Direct Submission
JOURNAL
Submitted (17-NOV-1998) Australian Equine Blood Typing Research
Laboratory, University of Queensland, St. Lucia, Brisbane,
Queensland 4072, Australia

REFERENCE
1 (bases 1 to 11406)
AUTHORS
Maeel,A.M., Brandon,R.B. and Bell,T.K.
TITLE
Nucleotide Sequence of the Equine Beta-lactoglobulin Gene
JOURNAL
Unpublished
AUTHORS
Maeel,A.M., Brandon,R.B. and Bell,T.K.
TITLE
Direct Submission
JOURNAL
Submitted (17-NOV-1998) Australian Equine Blood Typing Research
Laboratory, University of Queensland, St. Lucia, Brisbane,
Queensland 4072, Australia

FEATURES

source

1..11406

/organism="Equus caballus"

/mol_type="genomic DNA"

/db_xref="taxon:9796"

<4472..>8679

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

REFERENCE
1 (bases 1 to 11406)
AUTHORS
Maeel,A.M., Brandon,R.B. and Bell,T.K.
TITLE
Nucleotide Sequence of the Equine Beta-lactoglobulin Gene
JOURNAL
Unpublished
AUTHORS
Maeel,A.M., Brandon,R.B. and Bell,T.K.
TITLE
Direct Submission
JOURNAL
Submitted (17-NOV-1998) Australian Equine Blood Typing Research
Laboratory, University of Queensland, St. Lucia, Brisbane,
Queensland 4072, Australia

REFERENCE
1 (bases 1 to 11406)
AUTHORS
Maeel,A.M., Brandon,R.B. and Bell,T.K.
TITLE
Nucleotide Sequence of the Equine Beta-lactoglobulin Gene
JOURNAL
Unpublished
AUTHORS
Maeel,A.M., Brandon,R.B. and Bell,T.K.
TITLE
Direct Submission
JOURNAL
Submitted (17-NOV-1998) Australian Equine Blood Typing Research
Laboratory, University of Queensland, St. Lucia, Brisbane,
Queensland 4072, Australia

FEATURES

source

1..11406

/organism="Equus caballus"

/mol_type="genomic DNA"

/db_xref="taxon:9796"

<4472..>8679

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956..5095,5767..5840,7367..7480,

8302..8406,8663..8679)

/gene="BLG"

join(4472..4567,4956

once, or longer because we provide a small overlap between neighboring submissions.

FEATURES
source
Location/Qualifiers

1. 70387
/organism="Caenorhabditis briggsae"
/mol_type="genomic DNA"
/strain="Gujarat G16"
/db_xref="taxon:6238"
/clone="CB023K10"
23139. 23209
/product="tRNA-Gly"
/note="codon recognized: GGC"
complement(36569. 36640)
/product="tRNA-Glu"
/note="codon recognized: GAG"
37437. 37508
/product="tRNA-Gln"
/note="codon recognized: CAA"

ORIGIN

Query Match 64.7%; Score 19.4; DB 3; Length 70387;
Best Local Similarity 79.3%; Pred. No. 5.7e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 TGGACCCCTCGACTCGAGAGTTCCGCT 30
Db 6252 TGGCACCTCTTCGATTGAAAGTTGAGCT 6280

RESULT 12

AE016816 6/c

WPCOMMENT

Sequence split into 9 fragments LOCUS AE016816 Accession AE016816

Fragment Name Begin End
AE016816_0 1 110000
AE016816_1 100001 210000
AE016816_2 200001 310000
AE016816_3 300001 410000
AE016816_4 400001 510000
AE016816_5 500001 610000
AE016816_6 600001 710000
AE016816_7 700001 810000
AE016816_8 800001 907057
Continuation 7 of 9 of AE016816 from base 600001 (AE016816 Eremothecium goesypii ATCC

Query Match 64.7%; Score 19.4; DB 8; Length 110000;
Best Local Similarity 79.3%; Pred. No. 5.5e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CTGGACCCCTCGACTCGAGATTCCGC 29
Db 20976 CTGGACTGCTGTCGATCGAGACTTCAC 20948

RESULT 13

AC104791/c

LOCUS AC104791 159969 bp DNA linear PRI 21-FEB-2002
DEFINITION Homo sapiens BAC clone RP11-181K12 from 4, complete sequence.
ACCESSION AC104791 AC032008
VERSION AC104791.3 GI:18482313
KEYWORDS HTG.
SOURCE Homo sapiens (human).
ORGANISM Homo sapiens (human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS Sulston, J.E. and Watson, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 159969)
AUTHORS Isek, A., Meyer, R. and Creason, K.

TITLE The sequence of Homo sapiens BAC clone RP11-181K12
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 159969)
AUTHORS Watson, R.H.

TITLE Direct Submission
JOURNAL Submitted (21-DEC-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 4 (bases 1 to 159969)
AUTHORS Watson, R.H.

TITLE Direct Submission
JOURNAL Submitted (03-FEB-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 5 (bases 1 to 159969)
AUTHORS Watson, R.

TITLE Direct Submission
JOURNAL Submitted (21-FEB-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Feb 3, 2002 this sequence version replaced gi:18042374.

COMMENT

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: saplens@watsn.wustl.edu

----- Summary Statistics

Center project name: H_NH0181K12

Drafting Center: WIBR

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, B., Tateo, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-292M9; the clone sequenced to the right is RP11-203B7. Actual start of this clone is at base position 1 of RP11-181K12; actual end is at base position 159969 of RP11-181K12.

A transposon has been identified in the vector of this clone.

The sequence of AC032008 has been incorporated into AC104791.

FEATURES
source

1. 159969
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

```

/chromosome="4"
/map="4"
/clone="RP11-181K12"
/clone_lib="RPCT-11"
repeat_region
/rpt_family="MIR"
191..463
/rpt_family="MIR"
1855..2217
repeat_region
/rpt_family="MALR"
2333..2362
/rpt_family="AT_rich"
2471..2690
/rpt_family="MIR"
3143..3169
repeat_region
/rpt_family="(A)n"
3296..3316
repeat_region
/rpt_family="AT_rich"
3304..3462
/rpt_family="Alu"
3589..3609
repeat_region
/rpt_family="AT_rich"
3984..4033
repeat_region
/rpt_family="ERV_L"
4019..4111
repeat_region
/rpt_family="(TA)n"
4087..4495
/rpt_family="ERV_L"
5195..5483
repeat_region
/rpt_family="Alu"
5374..5402
repeat_region
/rpt_family="(GGAG)n"
5897..5927
repeat_region
/rpt_family="(TTTTG)n"
5906..6217
repeat_region
/rpt_family="Alu"
6577..6760
/misc_feature
/Note="Similar to Homo sapiens EST AW843469
(NID:g7937452)"
7129..7131
/misc_feature
/Note="Similar to Homo sapiens EST AW843469
(NID:g7937452)"
7314..7393
repeat_region
/rpt_family="MIR"
7394..7415
repeat_region
/rpt_family="(T)n"
7700..7766
/rpt_family="MIR"
9154..9438
repeat_region
/rpt_family="Alu"
9437..9466
repeat_region
/rpt_family="(TAA)n"
10074..10204
/rpt_family="MIR"
10351..10812
repeat_region
/rpt_family="MALR"
11128..11343
repeat_region
/rpt_family="Alu"
11414..11626
repeat_region
/rpt_family="MIR_type"
13354..13449
repeat_region
/rpt_family="MIR_type"
13450..13721
repeat_region
/rpt_family="Alu"
13722..13784
repeat_region
/rpt_family="MIR_type"
13813..13852
repeat_region
/rpt_family="CT-rich"
15896..15941
repeat_region
/rpt_family="(TG)n"
16404..16916
repeat_region
/rpt_family="L2"
16918..16939
repeat_region
/rpt_family="(T)n"
18098..18220

```

```

/rpt_family="L1"
18250..18539
/rpt_family="Alu"
18545..18594
/rpt_family="(A)n"
18829..19205
/misc_feature
/Note="match to EST AA743144 (NID:g2782650) my20h09.s1"
19565..19869
repeat_region
/rpt_family="Alu"
20046..20785
repeat_region
/rpt_family="ERV_L"
21452..21689
/rpt_family="L1"
21764..22061
repeat_region
/rpt_family="Alu"
21780..22172
/misc_feature
/Note="match to EST AA554485 (NID:g2325024) n136b03.s1"
22691..22990
repeat_region
/rpt_family="Alu"
23268..23312
repeat_region
/rpt_family="(TTTA)n"
23286..23596
/rpt_family="Alu"
23637..24312
/rpt_family="MIR2_type"
24463..24756
/rpt_family="L2"
24798..25023
/rpt_family="MIR2_type"
25050..25411
/rpt_family="ERV_L"
25377..25406
/rpt_family="(CAA)n"
25412..25733
/rpt_family="Alu"
25532..25561
/rpt_family="(A)n"

```

Query Match 64.7%; Score 19.4; DB 9; Length 159969;
 Best Local Similarity 79.3%; Pred. No. 5.3e+02;
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 TGGACCCCTCGACTCGAGAGTTCCGCT 30
 Db 120906 TGGACTACTCTGCTCCAGACTTCTCT 120878

RESULT 14
 AP006440/c 189269 bp DNA linear HTG 21-MAY-2003
 LOCUS Homo sapiens chromosome 11 clone RP11-368120 map 11q, WORKING DRAFT
 DEFINITION
 ACCESSION AP006440
 VERSION AP006440.1 GI:30962586
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE
 1 Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,
 Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 Homo sapiens genomic DNA of 11q
 Published Only in Database (2003)
 2 (bases 1 to 189269)
 Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,
 Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 Direct Submission
 Submitted (19-MAY-2003) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail: hattori@gsc.riken.go.jp URL: http://hgp.gsc.riken.go.jp/,
 Tel: 81-45-503-9111, Fax: 81-45-503-9170)

JOURNAL

COMMENT

----- Genome Center
Center: RIKEN Genomic Sciences Center (GSC)

Center code: RIKEN

Web site: <http://hgp.gsc.riken.go.jp/>
Contact: hattori@gsc.riken.go.jp

----- Project Information
Center project name: HumDrafc11

Center clone name: RP11-368120

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 188057 bases at least Q40

Consensus quality: 188482 bases at least Q30

Consensus quality: 188825 bases at least Q20

Insert size: 188969; sum-of-contigs

Quality coverage: 12.20x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs 'N', but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 70266 contig of 70266 bp in length
70367 130873 contig of 60507 bp in length
130974 188629 contig of 57656 bp in length
188730 189269 contig of 540 bp in length.
* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of 'N', but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
* 1 70266: contig of 70266 bp in length
* 70267 70366: gap of 100 bp
* 70367 130873: contig of 60507 bp in length
* 130874 130973: gap of 100 bp
* 130974 188629: contig of 57656 bp in length
* 188630 188729: gap of 100 bp
* 188730 189269: contig of 540 bp in length.
Location/Qualifiers
1. 189269
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-368120"
1. 70266
/note="assembly_fragment"
70367. 130873
/note="assembly_fragment clone_end:r7 vector_side:right"
130974. 188629
/note="assembly_fragment"
188730. 189269
/note="assembly_fragment clone_end:sp6 vector_side:right"

FEATURES

source

misc_feature
/note="assembly_fragment"
1. 70266
misc_feature
/note="assembly_fragment"
70367. 130873
misc_feature
/note="assembly_fragment clone_end:r7 vector_side:right"
130974. 188629
misc_feature
/note="assembly_fragment"
188730. 189269
/note="assembly_fragment clone_end:sp6 vector_side:right"

ORIGIN

Query Match 64.7%; Score 19.4; DB 2; Length 189269;
Best Local Similarity 79.3%; Pred. No. 5.3e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 TGAGCCCTCTGAGTCGAGAGTCCGCT 30
|||||
Db 161287 TGAGTACTCTCTCTCCAGACTTCTCT 161259

RESULT 15
AC011862/c

LOCUS

AC011862 196472 bp DNA linear HTG 13-JUL-2000
DEFINITION Homo sapiens clone RP11-15K8, LOW-PASS SEQUENCE SAMPLING.

AC011862.4 GI:7144913

HTG; HTGS_PHASE0.
KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 196472)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome, clone RP11-15K8

Unpublished

2 (bases 1 to 196472)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,

Baldwin, J., Barn, N., Beckerly, R., Boguslavsky, L., Boukhgalter, B.,

Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,

Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donegan, L., Doyle, M.,

Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,

Galagan, J., Gardyna, S., Grant, G., Hagos, B., Headford, A., Horton, L.,

Howland, J.C., Johnson, R., Jones, C., Kann, L., Karas, A., Klein, J.,

Lehoczky, J., Lien, C., Locke, K., MacDonald, P., Margulis, N.,

McEwan, P., McQuirk, A., McKernan, K., McLaughlin, D., Meldrum, J.,

Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,

Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,

Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,

Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,

Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (15-OCT-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 3, 2000 this sequence version replaced GI:6980935.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RW/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: 13484

Center clone name: 15_K_8

* NOTE: This record contains 217 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 883: contig of 883 bp in length
884 983: gap of 100 bp
884 1877: contig of 894 bp in length
884 1877: gap of 100 bp
1878 2866: contig of 889 bp in length
1978 2866: gap of 100 bp
2867 3829: contig of 863 bp in length
2867 3829: gap of 100 bp
3830 3929: gap of 100 bp
3830 4831: contig of 902 bp in length
4831 4931: gap of 100 bp
4832 4931: contig of 859 bp in length
4931 5790: gap of 100 bp
5790 6752: contig of 862 bp in length
5791 6752: gap of 100 bp
6753 7749: contig of 897 bp in length
6753 7749: gap of 100 bp
7750 7850 7849: gap of 100 bp
7850 8754: contig of 905 bp in length
8755 8854: gap of 100 bp

```
* 8855 9749: contig of 895 bp in length
* 9750 9849: gap of 100 bp
* 9850 10738: contig of 889 bp in length
* 10739 10838: gap of 100 bp
* 10839 11725: contig of 887 bp in length
* 11726 11825: gap of 100 bp
* 11826 12707: contig of 882 bp in length
* 12708 12807: gap of 100 bp
* 12808 13697: contig of 890 bp in length
* 13698 13797: gap of 100 bp
* 13798 14668: contig of 871 bp in length
* 14669 14768: gap of 100 bp
* 14769 15648: contig of 880 bp in length
* 15649 15748: gap of 100 bp
* 15749 16622: contig of 874 bp in length
* 16623 16722: gap of 100 bp
* 16723 17587: contig of 865 bp in length
* 17588 17687: gap of 100 bp
* 17688 18558: contig of 871 bp in length
* 18559 18658: gap of 100 bp
* 18659 19534: contig of 876 bp in length
* 19535 19634: gap of 100 bp
* 19635 20510: contig of 876 bp in length
* 20511 20610: gap of 100 bp
* 20611 21517: contig of 907 bp in length
* 21518 21617: gap of 100 bp
* 21618 22499: contig of 882 bp in length
* 22500 22599: gap of 100 bp
* 22600 23480: contig of 881 bp in length
* 23481 23580: gap of 100 bp
* 23581 24444: contig of 864 bp in length
* 24445 24544: gap of 100 bp
* 24545 25479: contig of 935 bp in length
* 25480 25579: gap of 100 bp
* 25580 26492: contig of 913 bp in length
* 26493 26592: gap of 100 bp
* 26593 27454: contig of 862 bp in length
* 27455 27554: gap of 100 bp
* 27555 28431: contig of 877 bp in length
* 28432 28531: gap of 100 bp
* 28532 29495: contig of 964 bp in length
* 29496 29595: gap of 100 bp
* 29596 30489: contig of 894 bp in length
* 30490 30589: gap of 100 bp
* 30590 31435: contig of 846 bp in length
* 31436 31535: gap of 100 bp
* 31536 32425: contig of 890 bp in length
* 32426 32525: gap of 100 bp
* 32526 33401: contig of 876 bp in length
* 33402 33501: gap of 100 bp
* 33502 34376: contig of 875 bp in length
* 34377 34476: gap of 100 bp
* 34477 35355: contig of 879 bp in length
* 35356 35455: gap of 100 bp
* 35456 36370: contig of 915 bp in length
* 36371 36470: gap of 100 bp
* 36471 37358: contig of 888 bp in length
* 37359 37458: gap of 100 bp
* 37459 38332: contig of 874 bp in length
* 38333 38432: gap of 100 bp
* 38433 39285: contig of 853 bp in length
* 39286 39385: gap of 100 bp
* 39386 40317: contig of 932 bp in length
* 40318 40417: gap of 100 bp
* 40418 41314: contig of 897 bp in length
* 41315 41414: gap of 100 bp
* 41415 42297: contig of 883 bp in length
* 42298 42397: gap of 100 bp
* 42398 43270: contig of 873 bp in length
* 43271 43370: gap of 100 bp
* 43371 44268: contig of 898 bp in length
* 44269 44368: gap of 100 bp
* 44369 45233: contig of 865 bp in length
```

```
* 45234 45333: gap of 100 bp
* 45334 46202: contig of 869 bp in length
* 46203 46302: gap of 100 bp
* 46303 47193: contig of 891 bp in length
* 47194 47293: gap of 100 bp
* 47294 48183: contig of 889 bp in length
* 48183 48282: gap of 100 bp
* 48283 49146: contig of 864 bp in length
* 49147 49246: gap of 100 bp
* 49247 50116: contig of 870 bp in length
* 50117 50216: gap of 100 bp
* 50217 51112: contig of 896 bp in length
* 51113 51212: gap of 100 bp
* 51213 52102: contig of 890 bp in length
* 52103 52202: gap of 100 bp
* 52203 53064: contig of 862 bp in length
* 53065 53164: gap of 100 bp
* 53165 54024: contig of 860 bp in length
* 54025 54124: gap of 100 bp
* 54125 54957: contig of 833 bp in length
* 54958 55057: gap of 100 bp
* 55058 55938: contig of 881 bp in length
* 55939 56038: gap of 100 bp
* 56039 56932: contig of 894 bp in length
* 56933 57032: gap of 100 bp
* 57033 57916: contig of 884 bp in length
* 57917 58016: gap of 100 bp
* 58017 58875: contig of 859 bp in length
* 58876 58975: gap of 100 bp
* 58976 59844: contig of 869 bp in length
* 59845 59944: gap of 100 bp
* 59945 60795: contig of 851 bp in length
* 60796 60895: gap of 100 bp
* 60896 61774: contig of 879 bp in length
* 61775 61874: gap of 100 bp
* 61875 62739: contig of 865 bp in length
* 62740 62839: gap of 100 bp
* 62840 63807: contig of 968 bp in length
* 63808 63907: gap of 100 bp
* 63908 64783: contig of 876 bp in length
* 64784 64883: gap of 100 bp
* 64884 65767: contig of 884 bp in length
* 65768 65867: gap of 100 bp
* 65868 66765: contig of 898 bp in length
* 66766 66865: gap of 100 bp
* 66866 67730: contig of 865 bp in length
* 67731 67830: gap of 100 bp
* 67831 68721: contig of 891 bp in length
* 68722 68821: gap of 100 bp
* 68822 69711: contig of 880 bp in length
* 69712 69811: gap of 100 bp
* 69812 70699: contig of 888 bp in length
* 70699 70799: gap of 100 bp
* 70800 71679: contig of 880 bp in length
```

```
Query Match 64.7% Score 19.4; DB 2; Length 196472;
Best Local Similarity 79.3%; Pred. No. 5.3e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 CTGGACCCCTCTGCACTCGAGAGTTCCGC 29
Db 6946 CTGGACCCCTGATGCTCGAGAGTTCCCC 6918
```

Search completed: March 1, 2005, 00:29:18
Job time : 883.065 secs

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☒ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

This Page Blank (uspto)